Variation in the Use and Norms of English Verb Forms:

A Study of Finnish University Students and Teachers

Satu von Boehm

Doctoral dissertation, to be presented for public discussion with the permission of the Faculty of Arts of the University of Helsinki, in the Great Hall of the Language Centre, on the 29th of May 2019 at 12 o'clock.

Language Centre Publications 8

Published by the University of Helsinki Language Centre
Fabianinkatu 26
FI-00014 University of Helsinki
Finland

COVER PHOTO Nina Dannert
COVER DESIGN Unigrafia

ISBN 978-951-51-5223-7 (nid.)
ISBN 978-951-51-5224-4 (PDF)

Printed by Unigrafia, Helsinki

This publication is part of a reviewed series designed to communicate and disseminate research and practice in foreign language education and related fields.

Previous publications in the series:

7. Karjalainen, S., Jokinen, J. & Mäkäläinen, H. (Eds.) 2018. *Kielenoppimisen kehittyvät arviointi- ja palautekäytänteet – Developing feedback and assessment practices in language learning*.
6. Lehtonen, T. & Vaattovaara, J. (Eds.) 2016. *Näkökulmia kielenoppimisen ohjaukseen – On advising and counselling in language learning*.
5. Bradley, F., Lappalainen, P. & Braidwood, E. (Eds.) 2015. *Papers from the 26th Communication Skills Workshop: Identity Explored: Language Centre, Language Professional, Language Teacher, Language Learner*.
4. Matilainen, M., Siddall, R. & Vaattovaara, J. (Eds.) 2013. *Muuttuva Kielikeskus – Språkcentrum i förändring – Language Centre in change*.
3. Korhonen, S. 2012. *Oppijoiden suomi: Koulutettujen aikuisten käsitykset ja kompetenssit (Perceptions and competences of adult learners of Finnish)*.
2. Pitkänen, K.K., Jokinen, J., Karjalainen, S., Karlsson, L., Lehtonen, T, Matilainen, M., Niedling, C. & Siddall, R. (Eds.) 2011. *Out-of-Classroom Language Learning*.
1. Karlsson, L. 2008. *Turning the Kaleidoscope. (E)FL Educational Experience and Inquiry as Auto/biography*.

HELSINGIN YLIOPISTON KIELIKESKUKSEN JULKAISUSARJA
PUBLIKATIONER UTGIVNA AV SPRÅKCENTRUM VID HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI LANGUAGE CENTRE PUBLICATIONS

ISSN 1797-9269

Abstract

This study focuses on variation in Finnish university students' use of English verb forms and teachers' assessment of such use. The goal of the study was to identify the nature and extent of variation in both students' responses and teachers' assessment of these responses. The study demonstrates that there is extensive variation in both students' use of English verb forms and in teachers' assessment of the forms.

The data consist of students' responses to a fill-in-the-gap test. The study comprised 319 students attending English courses at three Finnish universities in 2003-2004. In addition, 13 English language teachers at the University of Helsinki evaluated the level of acceptability of the students' responses. The study was conducted by drafting a list of all the suggested answers and asking teachers of English to rate the answers with a four-point scale. After this, the responses were analysed with different levels of strictness. The criteria for this depended on how many teachers were required to judge an answer successful.

The results indicate that the level of expected teacher consensus radically affected the results. The criteria applied in assessment have a critical impact on the impression created of the students' proficiency. The range of variation among teachers mainly resulted from their different interpretations of event time and their reactions to spelling errors and to providing unconventional tense or aspect. Finnish, American and British teachers of English had negligible differences at the group level: the differences were caused by the individual norms the teachers followed. Reliance on only one teacher's assessment of a student's performance may create arbitrary results, and more teachers are needed for reliable evaluation, at least for high-stakes purposes.

The results also indicate that although some students are skilful at providing the expected verb forms, many students struggle with the provision of any forms beyond the simple past. It seems that investment in communicative teaching practices, although beneficial for promoting students' fluency in English, does not assist them in their attempts at greater accuracy. Some further attention to grammar would provide students with more options to express their message accurately, proficiently, meaningfully and intelligibly.

Of the background factors, increased exposure to and use of English or additional study did not influence students' skills in the use of English verb forms. Regardingschool and matriculation examination marks, only high marks predicted good performance in the test, while marks in the middle and low end of the spectrum were not reliable indicators of students' ability to provide verb forms accurately.

The findings imply that researchers, raters and teachers need to become more aware of the limitations of reliance on intuition alone in evaluating students' language skills. Some parts of language use do not readily bend to existing norms. The degree to which teachers are likely to accept unexpected forms significantly contributes to how learners are treated in testing situations. While it is useful to attempt to harmonise testing practices, it is also important to acknowledge the fact that there is variation both in learners' use of English and in teachers' assessment of such use.

Abstract

Denna undersökning fokuserar på den variation som finns inom språkbruket i engelska bland finländska studerande och på deras kunskaper i att använda engelska verb. Studien fokuserar också på universitetslärares åsikt om hur korrekt studerandenas verbanvändning på engelska är. Målet är att undersöka hurdan variation det finns när det gäller verbanvändning och hur utbredd den är. Resultaten visar att variationen är stor både när det gäller studerandenas verbanvändning och när det gäller lärarnas bedömning av denna.

Materialet består av 319 studerandenas svar i ett lucktest. Dessa studerande deltog i en kurs i engelska vid tre högskolor i Finland läsåret 2003-2004. Dessutom utvärderade tretton språklärare vid Helsingfors universitet svaren utifrån hur acceptabla dessa var. Alla givna svar skrevs upp på en lista, och engelskalärarna bedömde alla svar på en fyrgradig skala. Efter detta analyserades svaren med olika nivåer av strikthet på basis av hur många lärare som var av samma åsikt om varje svars riktighet.

Resultaten pekar på att studerandenas nivå i engelska beror på hur man tolkar lärarnas bedömning. Hur korrekta svaren är beror på hur många lärare med samma åsikt som behövs för att definiera någon verbform som rätt svar. Grunden till den variation som förekommer bland lärarnas bedömning kan anses bero dels på att lärarna tolkade tidsaspektenen inom en viss kontext på olika sätt, på att de bedömt misstag i rättskrivning på olika sätt och dels på lärarnas olika tolkningar av vilken tidsform som ska användas. Undersökningen omfattade finländska, amerikanska och brittiska lärare. På gruppnivå var skillnaderna minimala, eftersom de flesta skillnader berodde på individuella normtolkningar och på lärarnas olika kännedom om alternativa former. Därför är det viktigt att komma ihåg att om man bara litar på en enda lärares bedömning av en studerande kan resultatet vara opålitligt, och om man testar för ändamål som kan ha viktiga konsekvenser, borde man använda flera testpersoner för att nå pålitliga resultat.

Resultaten visar också att även om en del studerande kunde använda många olika verbformer i engelskan, fanns det många studerande som bara kunde använda preteritumformen korrekt. Detta indikerar till att betoningen av kommunikativa färdigheter i språkundervisningen är nyttig för att höja studerandenas förmåga att uttrycka sig flytande på engelska, men detta verkar inte förbättra studerandenas kunskaper i grammatiken. Det vore därför bra att fokusera mera på vissa aspekter inom grammatiken för att studerandena ska lära sig att formulera det som de vill uttrycka på ett mer korrekt, mer förståeligt, mer flytande och mer noggrant sätt.

Av studerandernas bakgrundsinformation kan vi observera att ökade möjligheter att lyssna och läsa på engelska, att använda engelska och att studera extra kurser inte påverkar studerandenas färdighet i att använda verbformer i engelskan. När det gäller bedömningen i skolan och i studentexamen, är det bara de högsta vitsorden som kan förutse goda resultat i detta test, medan medelmåttiga eller sämre vitsord inte kunde användas för att förutse studerandenas förmåga att på ett ändamålsenligt sätt använda olika verbformer i engelskan.

Resultaten implicerar att forskare, lärare och testare bör vara medvetna om att de inte endast kan lita på sin intuition då de utvärderar studerandenas kunskaper. Vissa delar av språket är svårdefinierbara. Hur läraren reagerar på ovanliga former påverkar hur studerandena utvärderas i bedömningssituationer. Det är dock en bra idé att fortsätta harmonisera testmetoderna, men samtidigt är det också viktigt att vara medveten om att det finns variation i studerandenas språkproduktion och i lärarnas bedömning av sådan produktion.

Abstrakti

Tämä tutkimus käsittelee suomalaisten opiskelijoiden englannin kielen verbimuotojen käytössä esiintyvää variaatiota sekä variaatiota siinä, miten opettajat arvioivat näiden opiskelijoiden englannin kielen käyttöä. Tavoitteena on tarkastella, millaista variaatiota esiintyy ja kuinka laajamuotoista se on. Tutkimus osoittaa, että sekä opiskelijoiden tarjoamissa verbimuodoissa että opettajien arvioinnissa on merkittävää vaihtelua.

Tutkimusaineisto koostuu 319 opiskelijan vastauksista aukkotehtäviin. Opiskelijat kävivät englannin kurssia kolmessa suomalaisessa korkeakoulussa vuonna 2003-2004. Sen lisäksi kolmetoista Helsingin yliopiston englanninopettajaa arvioi opiskelijoiden vastausten hyväksyttävyyttä. Kaikki opiskelijoiden tarjoamat vastaukset kerättiin, ja osallistuvat opettajat arvoivat kunkin vastauksen neliportaisella asteikolla. Nämä vastaukset analysoitiin neljällä eri tasolla sen mukaan, kuinka monen opettajan oli oltava samaa mieltä opiskelijan vastauksen oikeellisuudesta.

Tulokset osoittavat, että opiskelijoiden suoritustason arviointi riippuu siitä, millä kriteereillä tuotosta arvoidaan ja kuinka monta opettajaa tarvitaan määrittelemään vastausten oikeellisuus. Erot opettajien arvioinnissa selittyvät pitkälti sillä, miten he suhtautuivat esimerkiksi oikeinkirjoitusvirheisiin, epätavanomaiseen aikamuotoon tai oletetusta poikkeavaan tulkintaan tapahtuma-ajasta. Tutkimukseen osallistui suomalaisia, amerikkalaisia ja brittiläisiä opettajia, ja ryhmätasolla heidän välisensä erot olivat minimaalisia; sen sijaan eroja aiheuttivat yksilölliset suhtautumistavat oletettujen normien rikkomiseen sekä tietoisuus vaihtoehtoisista muodoista. Tästä syystä onkin huomattava, että vain yhden opettajan tekemä arviointi voi johtaa sattumanvaraisiin tuloksiin ja että luotettavampiin tuloksiin arvioinnissa päästään käyttämällä useampia opettajia arvioijina. Tämä on erityisen tärkeää silloin, kun arvioinnilla on testattavalle laajakantoisia seurauksia.

Tulokset osoittavat myös, että vaikka osa opiskelijoita osaa käyttää englannin verbimuotoja taitavasti, jotkut osaavat käyttää pelkkää yleisimperfektimuotoa oikein. Näyttääkin siltä, että viimeaikainen kommunikatiivisen kielenopetuksen lisääntyminen on vaikuttanut suotuisasti suomalaisopiskelijoiden englannin kielen käytön sujuvuuteen mutta sillä ei ole ollut vaikutusta oikeakielisyyteen. Huomion kohdentaminen tiettyihin kieliopin kohtiin tarjoaisi opiskelijoille nykyistä paremmat mahdollisuudet viestiä tarkasti, sujuvasti, tarkoituksenmukaisesti ja ymmärrettävästi.

Opiskelijoiden taustatiedoista kävi ilmi, että lisääntynyt altistuminen englannin kielelle tai ylimääräiset kielikurssit eivät vaikuttaneet opiskelijoiden osaamistasoon. Koulun ja ylioppilastodistuksen arvosanoista pelkästään korkeimmat arvosanat ennustivat hyvää osaamistasoa, kun taas keskitason ja huonommat arvosanat eivät olleet luotettavia osoittamaan opiskelijoiden kykyä käyttää englannin verbejä tarkoituksenmukaisesti.

Tulosten perusteella tutkijoiden, opettajien ja kielitaidon arvioijien on hyvä tiedostaa, että pelkkä intuitio ei riitä arvioimaan opiskelijoiden taitotasoa. Jotkut kielen alueet ovat vaikeasti määriteltävissä, ja sillä, missä määrin opettaja suvaitsee epätavallisia muotoja, on suuri merkitys siihen, miten opiskelijan tuotos arvioidaan. Vaikka on hyödyllistä pyrkiä yhdenmukaistamaan testauskäytäntöjä, on myös huomioitava se, että sekä opiskelijoiden kielenkäyttötaidoissa että opettajien arviointikäytännöissä on vaihtelua.

Acknowledgements

No-one writes a doctoral dissertation alone, and several people are acknowledged for the help and support they have given me. First, my two supervisors, Professors Terttu Nevalainen and Minna Palander-Collin, have been extremely patient with me. I wish to use this opportunity to thank them very much for the advice and support they have given me over these many, sometimes difficult years. Words cannot truly convey my gratitude!

Professors Scott Jarvis and Riikka Alanen have been instrumental in improving the quality of this study. Their suggestions provided me with invaluable advice on how to revise parts of the text. Professors Paula Kalaja and Simon Borg have also given me important advice and criticism, which have been crucial in improving the study. I would like to thank them very much for the time they invested and comments they provided.

I would like to thank the students and the teachers who responded to the survey and Cambridge University Press for permission to reproduce the test that was used in the study. The Language Centre of the University of Helsinki has been generous with providing facilities and teaching reductions in order for me to focus on research. For this, I am grateful to the Director, Ulla-Kristiina Tuomi, and the former Director, Pirkko Forsman-Svensson.

A warm 'thank you' goes to the teachers at the University of Helsinki Language Centre who have taken part in numerous discussions regarding grammar. In particular, I would like to thank Kari K. Pitkänen, who has proactively devoted countless hours to commenting on my text, and Tuula Lehtonen and Leena Karlsson for their faithful support. I owe a debt of gratitude to Fergal Bradley, Deborah Clarke, Martin Darling, Paul Graves, Gráinne Hiney-Lehtipuu, Edward Hunter, Kenneth Kidd, Tuula Lehtonen, Jonathon Martin, Natalie Palmen, Kari K. Pitkänen, Roy Siddall, Kirby Vincent and Joseph Warnes for their help in proofreading the text. Furthermore, I thank all my colleagues at the English Unit for their help and interest over the years, and I am grateful to the members of the Writing Circle for the Thursday meetings.

I would like to thank Viira Kiviranta and Maria Mannil for helping me translate the students' questionnaire into Swedish, Åsa Mickwitz and Leena-Maija Åberg for comments and discussions on the use of Swedish and Kimmo Svinhufvud for comments and discussions related to Finnish. I also thank the Language Centre Research Group for their help with the final stages of the manuscript. I owe a debt of gratitude to Kaija Ervola, Leena Koskinen, Maire Mäkinen, Nanna Qvist, Jarno Raukko and Robin Voster for asking their students to take part in the survey. Parts of this study have been presented at several seminars and conferences, and I would like to thank the participants for their useful questions and comments.

Finally, of course, my family and friends have been extremely supportive. In particular, I want to thank my husband for making it possible for me to work on this project over these many, many years. My children have been patient with and interested in my work, and my parents-in-law have been invaluable in minding the children and giving me time to write. Thank you so much!

Contents

List of figures

List of tables

List of abbreviations

CEFR	Common European Framework of Reference for Languages
EFL	English as a foreign language
ELF	English as a lingua franca
ENL	English as a native language
ESL	English as a second language
L1	First language (native language, mother tongue)
L2	Second language

1 Introduction

This study focuses on variation in the use of English verb forms with two distinct but intertwined groups of people. On the one hand, I study variation in the use of English verb forms provided by Finnish university-level students in a specific fill-in-the-gap test, and on the other hand, I examine variation in university-level English teachers' assessment of these forms. The study includes an analysis of the proficiency level of Finnish students as evidenced in teachers' acceptance ratings of the forms the students provide. In this manner, it focuses on both variation and proficiency.

It comes as no surprise that there is variation in the use of English verb forms among both native speakers and learners of English. It is not surprising, either, that language teachers may react differently to some of these forms. What can be surprising, however, is that the differences in the reactions may be extreme: the best possible form for one teacher may be completely unacceptable for another – even in the same context. If teachers of English are not unanimous in determining what the expected form is, what types of verb forms and contexts cause the most disagreement, and what do the teachers agree on? On the other hand, how common is it for students to offer non-standard forms, and what are such forms like? These questions are related to the norms the teachers adopt in rating the students' answers. This study aims to contribute to the pool of research on teachers' acceptability ratings and variation in the use of verb forms, both in student responses and in teachers' assessment of such responses. It also discusses potential implications for teaching English in Finland.

The study can be located within the research traditions of variation in learner language as well as variation in language assessment, as the focus is on variability in learners' use of English as well as on variability in teachers' evaluation of learners' language use. The study contributes to the discussion on how teachers' assessment of learners' language use varies and the factors that account for this. The study also addresses the concept of error and highlights the fact that teachers may approach errors differently. Furthermore, it explores the extent of variation that students exhibit in providing both expected and unconventional forms in a given context depending on their interpretation of the context and their language skills.

1.1 Personal rationale

My interest in this topic stems from a variety of directions. I became interested in variation in grammar when teaching a Remedial Grammar course at the Language Centre of the University of Helsinki for the first time in 2001. I became fascinated with the variation in the forms suggested by students to a particular fill-in-the-gap exercise, which I also used in this study (see Section 6.1). At times, the forms some students offered did not make much sense in the given context, and it remained a mystery why some students had difficulty with such verb forms that are used fairly similarly in English and Finnish (see Section 5.1). Another source of inspiration was a comment given by a teacher of another language at the Language Centre. According to her, everyone entering the university already had a mastery of English, and consequently, there was no need to teach any more English. I wanted to investigate whether this was really the case, as my experience proved otherwise.

An aspect I soon decided to include in the study was fellow teachers. I wanted to explore whether teachers agree on which forms are entirely inappropriate and which are the best ones for particular contexts. As students meet a number of teachers during their path from primary school to secondary school and to university, the views teachers have on accuracy affect how students view their own skills and whether they feel pressure to study one particular grammar topic more extensively than another. I wanted to assess how widespread the potential disagreement on acceptable forms is and what might affect such variation. Are there any particular factors that make English teachers prefer one form to another, or does the variation simply depend on some teachers applying stricter rules than others? In what ways are such rules stricter? Furthermore, if a teacher is lenient, can the limits of this leniency be somehow identified?

Ultimately, however, my interest in the topic and the starting point for this study was pedagogical. As McKay (2006, 1) writes, "[f]or teachers, a primary reason for doing research is to become more effective teachers. Research contributes to more effective teaching, not by offering definitive answers to pedagogical questions, but rather by providing new insights into the teaching and learning process." Indeed, this has driven me forward: an attempt to understand how I can help students who are puzzled with the variety of forms from which to choose the appropriate verb form for a particular context.

1.2 Theoretical rationale

It is important to research variation in language use, and particularly the approach teachers take to evaluating learners' use of English, because the precise nature and identification of errors has not been discussed sufficiently in the existing literature. In particular, it is crucial to examine what constitutes an error and what criteria are used to either define or correct errors, as both researchers and teachers have often tended to rate learners' language skills in terms of intuition, which may not be normative at all. Teachers and researchers alike readily presuppose that their own assumptions of language use are the standard, but this may not be the case, and in some areas of grammar, universal norms may not even exist.

Errors have often been the focus of research in second language acquisition, but many researchers have ignored the possibility that the view they or the raters have had on the expected answers or the criteria they have applied may have been subjective. Thus, learners may have been treated differently depending on whose norms the researchers have followed and whose assessment they have relied on. It is important to further examine the criteria raters apply in their assessment of learner production to ensure that the findings are aligned and reliable. The expected correct alternative may not always be the only possible correct or acceptable alternative in a particular context, which is a topic in second language acquisition research that deserves greater attention[1].

The above is not to say that there has not been research on errors or on Finnish students' skills in English. There has, of course, been previous research on learners' acquisition of English verb forms and on Finnish students' English skills, but there are gaps in our understanding of the extent and nature of variation in these. In more specific terms, there is a need to combine variation in students' responses with teachers' reactions to them, and there is also a need to gain more information about Finnish university-level students' skills in the use of English verbs.

Previous research on Finnish students (see Section 5.4) has not focused on verb forms in particular. Many of the 20th century studies focus on crosslinguistic influence and often target the difference between having Finnish or Swedish as the first language (e.g. Jarvis 2000; Ringbom 1983; Sjöholm 1983), while some studies discuss students' overall skills in English (e.g. Sartoneva 1998; Takala 2004; Tuokko 2000). Often, these studies focus on pupils and

[1] I would like to thank Scott Jarvis for useful comments on this section in particular.

students in primary and secondary education (e.g. Lehtonen and Sajavaara 1985; Ringbom 1977; 1983; Sartoneva 1998; Takala 2004; Tuokko 2000), while there is less research on Finnish university students' skills in English. More recently, 21st century studies have often focused on people's own perceptions of what they can do in English and how communicative they are, including the views they have on the English skills needed, for example, at work or in their personal or social life (e.g. Leppänen and Nikula 2008; Leppänen et al. 2011). While Meriläinen (2010a; 2010b) addresses the way changes in curricula towards more communicative teaching practices have affected Finns' skills in English grammar and vocabulary, her data mainly focus on contrastive differences between Finnish- and Swedish-speaking students.

Although there has been a wealth of studies focusing on verb usage internationally at different levels, there is a need for more research on university-level students' use of English verbs. The previous studies on verb usage have often targeted specific forms and the mastery of these forms among a particular student population (see Chapter 4), or they have focused on learners' speed of judging a sentence to be either acceptable or inappropriate in English (see Section 3.3). Many of the studies on learner English have focused on students with a Germanic or Romance first language, or on Asian learners. Thus, there is a need for more research on learners with a Finno-Ugric language background.

Again, while there has been research on teacher assessment, studies on variation in teachers' assessment of specific features in English are scarce, as is research on Finnish teachers of English. In a number of studies, variation among teachers has been seen as a negative phenomenon that should be overcome through harmonisation attempts. Although some studies (e.g. Huhta et al. 2014; Kondo-Brown 2002; McNamara 2000) have acknowledged the fact that teachers' assessment varies, there is a need for more studies that focus on exploring the extent of variation in teachers' responses and attempt to describe the variation rather than condemn it. Furthermore, there is a need for more studies that employ several teachers to assess the students' responses instead of relying on only one teacher.

1.3 Focus and aims

This study focuses on the command of English verb forms among university-level students in the metropolitan area of Helsinki and the variation found in the forms they produce in a fill-in-

the-gap test (see Section 6.1). This study also focuses on the variation evident in teachers' responses to the variation within the students' responses. The students come from three different universities, study a variety of academic fields, speak either Finnish or Swedish as their first language and have studied English since primary school (see Section 6.2). All the teachers recruited for evaluating the students' responses worked at the University of Helsinki at the time of participation. Both Finnish and native teachers of English were included (see Section 6.3).

The aim of the study was to explore the nature and extent of variation within the student groups and to investigate what might account for some of the variation. Another aim was to explore the nature and extent of teacher variation and to identify what types of verb forms and contexts create consensus as well as what the teachers disagree on. This study had a practical and a pedagogical aim as well: the results can assist teachers in making informed decisions about the verb forms that they should focus on when teaching English grammar. Furthermore, the findings contribute to the pool of information we have about Finnish university students' language skills.

When choosing the appropriate research method, Dörnyei (2007, 307) argues in favour of a pragmatic (practical) approach (see also Gorard 2010; Johnson and Onwuegbuzie 2004; Phakiti and Paltridge 2015), and I have followed this advice. Methodologically, this study represents both basic and applied research (e.g. Oliver 2010; Phakiti and Paltridge 2015) using primary data (e.g. Brown and Rodgers 2002; McKay 2006, 5) in order to understand phenomena related to applied linguistics and second language acquisition. The study is also informed by research on language testing (e.g. Read 2015). This study mainly consists of quantitative research. I am interested in general trends and variables that can be used to account for group-level variation as well as deviant cases and variation at the individual level. The core research questions combine my own interests and gaps in previous research and are the following:

1. *What is the extent of variation in Finnish university students' use of English verbs in a fill-in-the-gap test, and what accounts for this variation?*
2. *What is the extent of variation in teachers' responses to the variation displayed by Finnish university students in their use of English verbs, and what accounts for this variation?*

These questions are complemented with a focus on the background factors that best account for success in the test and a focus on the nature of answers that provoke the most variation. This

includes factors such as gender, previous school and matriculation examination marks, additional study in English and the command of other languages.

1.4 Structure of the study

The discussion of previous research in Chapters 2 to 5 takes an aspect-by-aspect approach: a variety of aspects relevant to the topic are discussed (e.g. Paltridge and Phakiti 2015, 262). The relevant fields for this study include variation studies, studies on teaching and learning grammar in a second language and studies relating to the use of English verb forms. Furthermore, it is important to provide some information on the role of English in Finnish schools as well as on the difficulties Finns may have when learning English.

Chapter 2 opens with a discussion of variation in English. I introduce some challenges with the concept of a native speaker and contemporary attitudes related to non-native Englishes and to English as a Lingua Franca (ELF). I also focus on variation in learner language and outline why certain structures may pose more difficulties than others in learning the grammar of a second language. This chapter is important because it places the participants in my study into the broader picture.

I explore the position of grammar in language teaching in Chapter 3. I begin with a discussion on the role of teacher beliefs and attitudes in teaching and learning grammar and explore how learner errors have been treated in language teaching and testing. I also introduce grammaticality and acceptability judgement studies. Chapter 3 helps to understand the importance of teachers' and students' views on grammar, the role that grammar has in contemporary language education and some challenges in teaching and testing grammar.

In Chapter 4, I summarise previous studies on the acquisition of English verb forms. I focus on tense and aspect as well as verb-related challenges in academic writing. Finally, I explore some theories that argue for a learning hierarchy in the acquisition of verb forms. This provides a background to what is known about the acquisition and use of verb forms with people from various linguistic backgrounds.

Chapter 5 discusses the role of English in Finland. It highlights the main differences in verb use between English, Finnish and Swedish, which are the languages that all the students in this

study had studied at school. I also chart the position of English in Finland and introduce the curricula applied in Finnish schools. Finally, I examine what previous research has reported on Finnish students' command of English and their skills in national and international comparisons. This chapter helps to understand the educational and linguistic context that the students and some of the teachers in this study shared.

The empirical part begins with Chapter 6, where I provide details about the test I used and the testing context. I also present the student and teacher participants, including details about their background. Finally, I introduce the methods that I used and some results from pilot studies.

Chapter 7 provides the results on student proficiency. The chapter begins with general trends, after which the results are discussed in greater detail. I also provide information on how the students' background is reflected in their overall command of English verb forms and compare student groups with one another. This chapter explores the effect of different levels of strictness or leniency on students' success rates.

In Chapter 8, variation in particular features of verb use is discussed. I provide details on the results regarding tense, aspect and reported speech, after which I discuss some interesting phenomena that arise from the analysis, including the extent of the provision of unusual forms and how consistent students were in their responses. I also compare different student groups with one another.

Chapter 9, in turn, focuses on variation in teachers' responses. After a general discussion, I comment on the lack of consensus arising from only one teacher, after which I explore more extensive disagreement among teachers. Finally, I examine the relationship between teachers' assessment and their self-reported approaches to English grammar.

Chapter 10 discusses the extent and nature of variation in verb forms within the student and teacher population. This chapter discusses the students' overall ability, the teachers' views on the students' success in the test and the variation that the teachers exhibited in their responses.

Finally, Chapter 11 provides the conclusions, implications for teaching and studying English and a discussion of the limitations that the choices I have made may pose on the reliability or validity of the study. The study ends with recommendations for further research and closing remarks.

2 Variation and norms in English

All languages display variation; the English language is no exception. Variation is unavoidable and a natural consequence of the fact that languages are used by humans. There is at least regional, social, personal, historical, cultural, situational and register variation (e.g. Yule 2014; for an extensive discussion, see Crystal 2003). In addition to variation among native speakers, learners of a particular language display variation when using it (e.g. Ellis and Barkhuizen 2005; Regan 2013). Variation "refers to differences in linguistic form, or to two or more ways of saying the same thing" (Regan 2013, 272). People may have attitudes towards specific types of variation, beliefs about the amount and nature of variation that should be tolerated in a language and opinions on what deviations from standard forms might be a sign of.

This chapter discusses manifestations of variation in English at the general level, with a focus on both native and non-native English. Such a discussion is needed to understand the nature of variation from the perspective of native speakers, teachers and learners and to understand both the educational and real-life settings that the students and teachers in my study (see Sections 6.2 and 6.3) have been raised and continue to take part in. For a recent overview on research on variation, see Regan (2013), who argues that most variation is, actually, systematic. For a discussion on variation in second language acquisition, see Romaine (2003), and for a discussion on near-nativeness, see Sorace (2003).

Section 2.1 discusses contemporary views on the concept of a native speaker and different understandings of what Standard English actually entails. Section 2.2 focuses on learner English, English as a lingua franca and attitudes towards non-native English. Section 2.3 introduces studies that have addressed variation in learner English and the origins of such variation as well as some factors that affect the acquisition of grammar in a second language.

2.1 Native speakers and Standard English

Native speakers have traditionally been referred to as the norm when teaching English to students of other languages. Previously, the two norms that were considered the most relevant in teaching ESL (English as a second language) or EFL (English as a foreign language)[2] were British English and American English. Today, with a growing number of varieties of English both as a first and a second language, the situation has changed so that there are more speakers of ESL, EFL or ELF[3] (see Section 2.2) than 'native' speakers of English (e.g. Crystal 2003, 106-109; Mauranen, Hynninen and Ranta 2010, 183), and at times, it may be difficult to say what language(s) a person is a native speaker of (e.g. Brown and Larson-Hall 2012). This results in even wider variation and new types of problems. What is the norm that learners should now refer to and who can be considered a native speaker of English? If there are different norms, who can decide which of them learners should follow?

There are different understandings on how one becomes a native speaker and what that implies. For example, Hackert (2009) explores the concept of the English native speaker from a historical perspective. She argues that the term "conjures up a sense of being born to a speech community" so that the term appears to make language "inherited and non-negotiable" (Hackert 2009, 306). The native speaker as "an inviolate linguistic standard" is contrasted with the learner, "whose efforts are viewed as 'wrong' and in need of correction" (Butcher 2005, 15). In the 19th and early 20th century, nationalism played an important role in identifying and glorifying native speakers, and nations and languages were seen as one. The concept of a native speaker provided a way of drawing boundaries between various speaker groups (Hackert 2009, 315). Historically speaking, the term 'native' has many uses that are offensive, depreciative or "smack of colonialism and/or nationalism" (Butcher 2005, 16). Many British and American writers considered English the 'natural' world language because of colonisation and the increasing political, economic and social dominance of Great Britain and the United States.

[2] Although the two are often used interchangeably, originally the *second* language was learned in a country where the target language is spoken as the first language, while the *foreign* language was learned in a country where the target language was not in wide use. Nowadays, the term ESL is often used for both meanings.

[3] A lingua franca is a language that serves as a common language between speakers of different first languages. For a brief discussion, see Section 2.2 and for various definitions, see e.g. Jenkins, Cogo and Dewey 2011.

However, what is problematic in this view is that if "mother-tongue English is described as belonging to certain parts of the world only, speakers from other geographical, national, and racial backgrounds are automatically denied native-speaker status, no matter what their actual linguistic competence or proficiency is actually like" (Hackert 2009, 314). Since there are different socioeconomic settings in different English-speaking countries, with distinct differences in the way these people use English, it becomes difficult to say which of these varieties learners should attempt to study. Overall, there is increasing dissatisfaction with holding native speakers as the standard or as a point of reference for language learners (Hackert 2009, 305-306; see also Cook 2016; Ortega 2009 and Section 2.2). For a defence of the native speaker norm, see Lardiere (2013).

Furthermore, a person can be a native speaker from different perspectives or different levels. Mauranen, Hynninen and Ranta (2010, 184) define the term 'native speaker' as "someone who has acquired a language in childhood as the first language or as one of the first languages, and considers him- or herself a native speaker". Despite this, they suggest that in academic discourse, the term 'educated native speaker' should be abandoned as a role model because there are, actually, "no native speakers of *academic* English" (Mauranen, Hynninen and Ranta 2010, 184, emphasis original), and although having learned English in childhood offers an advantage for native speakers to learn the conventions of academic English (see also Section 4.3), there are many other qualities that also play a role.

Because of these issues, Butcher (2005, 13-23) calls for new, more inclusive terminology and argues that the assumptions of superiority allocated to native speakers have no place in modern society. Many researchers have started using the terms L1 (first language) speaker and L2 (second language) speaker (or user; see Section 2.2) instead of referring to native speakers and learners. This view focuses on the chronological order in which the languages were learned, acknowledging the fact that a person can even have several L1s and L2s, L3s, L4s and so on[4] (cf. Ortega 2009, 5; for a discussion on the terminology, see Rothman, Cabrelli Amaro and de Bot 2013). Figure 1 illustrates the relationship between some of the terms used in research. In this study, I use the terms L1 speaker and native speaker as well as L2 speaker, user and learner interchangeably (see also Section 2.2).

[4] Because it is sometimes difficult to know how many languages a particular person has already acquired, the concept of L2 is typically used to cover for L2, L3, L4 and Ln, particularly when reporting on group results, where the L2 is truly the L2 for some of the learners but the L3, L4 and so on for some others. For criticism against this practice, see Rothman, Cabrelli Amaro and de Bot (2013).

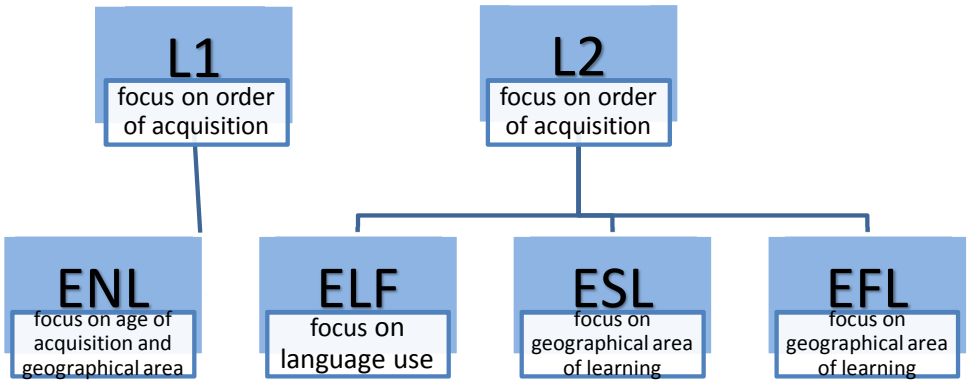


Figure 1. Terms used to refer to people using and learning English

While the insistence on using the native speaker norm and drawing on native usage is problematic, the term 'standard usage' is problematic as well, since there are actually several Standard Englishes. Indeed, Rhys (2007, 190) believes that Standard English only exists as an abstraction, not as an actual variety. Such a standard variety could only exist as a regional variety, and thus, varieties such as Standard British English or Standard American English may exist and be referred to as 'standard' (Rhys 2007, 190; Yule 2014, 243). According to Crystal (2003, 110), the Standard English of an English-speaking country is "a minority variety (identified chiefly by its vocabulary, grammar, and orthography) which carries most prestige and is most widely understood"; there is no World Standard English. Furthermore, changes in language use affect what is considered standard use (for a discussion on recent developments, see Leech et al. 2009 and Rhys 2007). However, Cook (2011, 147) maintains that despite the overwhelming variation in spoken English, "there is a global consensus with limited variation" in writing, spelling in particular. Even here, however, there is variability (e.g. Celce-Murcia and Larsen-Freeman 1999, 9; for variation in spoken grammar, see McCarthy and Carter 2002, 69-71).

In the empirical part of this study, I use university teachers of English whose first language is either English or Finnish as a point of reference for acceptable variation (see Section 6.3). For the purposes of this study, standard usage is discussed in terms of these teachers' acceptance ratings in addition to the suggested 'correct' forms provided in the answer key to the exercise book that the test is from (see Section 6.1). For the application of the term 'standard' in my study, see Chapters 7 and 8.

2.2 Reactions to non-native English

This section addresses the role of learner English and ELF in the world today by discussing reactions to non-native uses of English and to non-native speakers of English. The discussion includes the voice of the various stakeholders: learners, teachers, native speakers, non-native speakers, and the society at large. Such a discussion is needed because whether non-native speakers are treated as users or learners influences the way teachers and native speakers approach non-native English and the extent to which they tolerate variation. This also affects opinions on who is an acceptable role model and who is qualified to teach English.

Jenkins (2009, 202-203) offers a starting point for a distinction between ELF, which is used in intercultural communication primarily among speakers who are non-native speakers of English, and EFL, where the intention is to study English mainly for communication with native speakers. However, in today's world, most EFL speakers also aim at communication with other non-native speakers. Some scholars argue that 'ELFers' should be called 'users' rather than 'learners' of English, because the term 'user' better recognises the fact that "they are achieving a state of their own rather than perpetually trying to achieve an unattainable native speaker goal" (Cook 2011, 152; see also Jenkins 2009; Ortega 2009 and Section 2.1). Although the participants in my study can be conceived of as ELF users, I refer to them as learners because the test they completed was conducted in a learning environment[5] (see Section 6.1).

ELF is English that is used as "the common language of choice, among speakers from different linguacultural backgrounds" (Jenkins 2009, 200), irrespective of the geographical area where the speakers are located. ELF speakers need to make some adjustments to how they speak in order to be understood by their counterparts (e.g. Sewell 2010). ELF has some features that are found in ENL (English as a native language), but it also differs from ENL through the speakers' first languages and the influence of ELF contact (Jenkins 2009, 201). ELF interaction can include native speakers, but their role is not to provide linguistic models or norms. Although some items that differ from ENL are acceptable in ELF, this does not mean that 'anything goes' or that there are no incompetent ELF speakers: communicative effectiveness plays a key role in determining what is acceptable in ELF.

[5]In Kalaja's (2015a, 126-127) study, many pre-service Finnish teachers of English self-report that they had been taught grammar and vocabulary at school in the role of a learner and they had been English users mainly outside school.

Opinions vary on the role and acceptance of non-native English. In some studies, negative prejudice has been found towards non-native speakers of English (e.g. Birrell 2006; Lindemann 2002; Shibata 2010). Some native speakers do not seem to appreciate non-native speakers and may become uncooperative with them, failing to recognise that it takes two to reach mutual understanding (Lindemann 2002, 420). Furthermore, some studies indicate that native speakers may take a superior attitude and show disinterest towards communication with learners (Demont-Heinrich 2010; Friedrich 2010). This might also depend on how used these people are to communicating with non-native speakers.

A large number of universities in non-English-speaking countries now use English as the language of instruction. Some scholars (e.g. Benzie 2010; Birrell 2006) report concern over the "falling standards of learning on account of poor language skills, the status of local languages, and even falling standards of English" (Mauranen, Hynninen and Ranta 2010, 187) at universities around the world. Such worries might not be warranted, however. For example, studies carried out within the English as a Lingua Franca in Academic Settings project in Finland have shown that ELF interaction causes very little misunderstanding (Mauranen, Hynninen and Ranta 2010, 188) and that students interact both purposefully and successfully in ELF, although the linguistic patterns do not necessarily conform to native-speaker preferences. Despite this, some studies argue that increased exposure to English even at a university is not sufficient for learners to gain full command of English, but most learners also need high-quality language instruction in order to succeed (Benzie 2010; see also EF 2012). Staying in an English-speaking country does not automatically improve learners' language skills, nor does exposure to English without further attention paid to relevant aspects of language. Many researchers agree that some elements of grammar need to be "noticed" consciously (e.g. Andrews 2007, 15; Bardovi-Harlig and Dörnyei 1998; Ellis 2002a; Larsen-Freeman 2009, 527; Thornbury 1999, 24; see Chapter 3).

Some people, laymen in particular, believe that any native speaker is suitable as a role model or teacher of English, irrespective of their level of education or experience in teaching, and the status of non-native English teachers has been questioned. For example, Shibata (2010, 125) studied how Japanese high-school teachers of English perceive native and non-native assistant teachers of English and reports that junior high school teachers approved of both speakers, but senior high school teachers favoured native-speakers of English as the "ideal models". Furthermore, the teachers preferred non-native assistant teachers with "native-level grammar yet English recognisably accented by their native language" to those with "native-like

pronunciation yet minor inappropriate grammar use" (Shibata 2010, 126; cf. Section 3.1.2). Shibata believes that if the assistant teachers' role is to emphasise internationalisation, non-native teachers should be favoured in order to provide opportunities "to negotiate for intelligibility"; if the role is to serve as role models, then native speakers should be preferred (Shibata 2010, 133). This view, however, implies that such learners are studying English to become as close to native speakers as possible and not ELF users.

As ENL is often the model provided in textbooks, some ELF speakers consider ENL models superior despite the fact that their interlocutors are often not ENL speakers. For example, the non-native learner and teacher participants in Jenkins's study (2009, 204) favoured British and American English accents, both for correctness and pleasantness as well as for acceptability for international communication. She thinks this is surprising, as there is evidence that these accents are "not the most easily intelligible in lingua franca contexts" (Jenkins 2009, 204). Furthermore, respondents seemed to favour such non-native accents that were close to British or American accents and to appreciate ELF at the same time (Jenkins 2009, 204-205). Mauranen, Hynninen and Ranta (2010, 189) believe that, over time, the ENL speaker model is likely to have less dominance, while "the influence of professional and disciplinary communities" may increase.

In contrast to the views presented above, H. D. Brown (2007b, 137) suggests that since multiple varieties are now considered to be both legitimate and acceptable (see Section 2.1), teachers who have acquired English in an instructional setting have an advantage over native speakers precisely because of this experience. Nevertheless, although being a 'native speaker' does not automatically make one a good teacher (Butcher 2005, 15), many institutions recruiting language teachers only hire native speakers.

Could any variety of English, then, be taught in the classroom? In a very controversial study, Kirkpatrick, Deterding and Wong (2008, 374) report that "non-native speakers remain very cautious about accepting a non-native speaker model for the classroom" and argue that this would be because people fear that new varieties of English "may be unintelligible in international contexts" (Kirkpatrick, Deterding and Wong 2008, 359). Sewell (2010, 266), however, criticises their study for claiming that some native varieties of English are less intelligible than others, and explains that intelligibility is a feature of speakers and their linguistic forms, not something that a particular variety possesses. In his opinion, the world should accept the teaching of any variety that "can be shown to be internationally intelligible,

assuming of course that intelligibility, rather than identity affirmation, is an aim of the speaker" (Sewell 2010, 266). A further analysis of what intelligibility means would help teachers produce suitable learning materials for their local contexts.

Given the variation in English as an L1, what should be taught to learners of English? Cook (2011, 147) argues that "there is no question of teaching anything but the standard prestige version", despite the fact that it may be difficult to determine what that is (see Section 2.1). This means that learners are taught an "idealised abstract entity" (Cook 2011, 148). Nonetheless, Ur (2011, 508) argues that given the extent of ELF today, teachers should treat forms which seem acceptable in ELF contexts with increased leniency, but they should not begin to teach such forms. Thus, new skills and ways of communication are required among both native and non-native speakers of English, since for most L2 learners, the target is to use English for their personal and professional purposes (Cook 2011, 152). Cook (2016; 28) argues that L2 users should not be perceived as deficient speakers; rather, it needs to be acknowledged that they have "an independent language system of their own". In his opinion, comparing L2 users to native speakers is the same as comparing apples to oranges (Cook 2016, 28; see also Kalaja 2015b; Keck and Kim 2014; Ortega 2009).

As we saw above, both native and non-native speakers and learners of English may have prejudice towards particular varieties of English as well as towards speakers of learner English. However, students may be learning English for different purposes, and whether a teacher or an interlocutor treats the non-native user of English as a second or foreign language learner or as an ELF user affects how variation in their language use is perceived.

In summary, although there is prejudice over particular varieties and non-native speakers of English, the success or failure of communication does not depend on one speaker but the interlocutors' joint willingness to engage in communication in an attempt to understand one another. Furthermore, native speakers are not all alike, and in addition to geographical variation, there is sociocultural and socioeconomic variation. Given the variation in native English, it is difficult to identify one variety or accent that would serve as the model for all learners and for all the different needs they have. In order to help learners, language users and native speakers find mutual intelligibility, perhaps teaching materials should incorporate more examples of both spoken and written non-native English and of successful communication situations with ELF. This study contributes to this need through researching the extent of variation tolerance that teachers exhibit (see Chapter 9).

2.3 Variation in learner language

Obviously, learners' production of an L2 displays variation. While some of the variation is random or idiolectal, some of it is systematic either within an individual or between individuals (Regan 2013). In second language acquisition, the focus of attention has often been either on group patterns, with a quantitative approach, or on individual learning paths, with a qualitative focus. Importantly, it has been found that most group level phenomena hold at the individual level: Bayley and Langman (2004, 315; see also Regan 2013) conclude that we can often trace an individual learner's path from what research on groups reveals and make assumptions on the learner's success and problem areas. However, Regan (2004, 336-345), who followed second language learners over several years to see how group and individual patterns of variation interact, claims that although researchers tend to treat L2 learners as a homogeneous group, there is wide variety in individual differences. Despite the variation, her study also indicates that group results hold for individual variation. In many cases, variation is not a question of 'either/or' but, rather, 'more or less' (Regan 2013): users have a tendency to favour an alternative more than another, but typically employ both to some extent.

Because my study explores variation both at the group level and the individual level (see Chapters 7 and 8), this section presents some of the research that has been conducted on variation in learner language and the background factors that may predict or account for the variation. However, the reasons for the variation are wide and complex, and I can only capture a small proportion of research on this. Variation in second language acquisition is discussed extensively in e.g. Benati and Angelovska (2016), Gass and Selinker (2001), Lightbown and Spada (2006), Romaine (2003) and Saville-Troike (2012). Figure 2 illustrates some of the factors that have traditionally been considered to contribute to an individual's acquisition of a second language. For an extensive overview of such factors, see Dewaele (2013), Dörnyei and Ryan (2015), Dörnyei and Skehan (2003) and Jarvis and Pavlenko (2010); for some criticism, see Brown and Larson-Hall 2012 and Cook 2001.

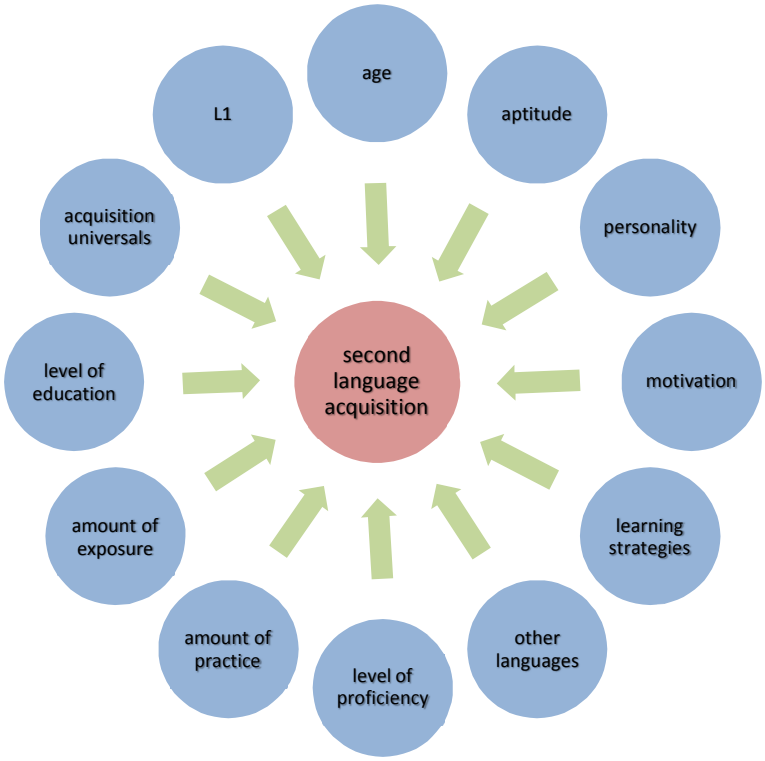


Figure 2. Factors that influence an individual's success in learning a second language

In this section, I focus primarily on the effect of the first language on the learning of subsequent languages. This is because this study focuses more on students' experience with languages and less on psychological factors such as personality or learning strategies (see Chapter 6). I begin with a brief discussion of how the first language influences the learning of other languages (Section 2.3.1) and continue with a discussion of factors that either assist or inhibit this process (Section 2.3.2). For a comprehensive discussion of such phenomena, see e.g. H. D. Brown (2007a; 2007b), Gass and Selinker (2001), Jarvis and Pavlenko (2010) and Ortega (2009).

2.3.1 The first language vs. additional languages

There is a multitude of studies, many of them partially conflicting, on the role of the L1 in second language acquisition (for a comprehensive overview, see Jarvis 2000, Jarvis and Pavlenko 2010 and Odlin 2003; for a historical overview, see Foley and Flynn 2013 and Thomas 2013). One of the reasons for the contradictory findings is variation in methodology, and Jarvis (2000, 248-249) calls for more uniform standards to ensure better reliability as there

is a lack of consensus concerning what to research when investigating crosslinguistic influence. L1 influence is the easiest to notice "when learners from different L1 backgrounds behave differently when using the same L2" (Jarvis 2000, 251), although this effect is usually evident in tendencies and in what learners are likely to do, not in stable patterns. Nevertheless, in many cases it is not L1 alone that influences the learning process (see Figure 2 above). Dependence on the L1 decreases over time, as learners modify their understanding of how the target language works and similarities within the target language are increasingly taken advantage of (Ringbom and Jarvis 2009, 114).

Although research on second language learning has often focused on differences between the L1 and the target language, learners actually seek for similarities rather than differences in order to learn the new language (Ringbom and Jarvis 2009, 106). When learners have little knowledge of the target language, their L1 is the main source for seeking similarities, but in case learners have already acquired some other languages, they can benefit from these languages as well, especially if they have been acquired to a high level of proficiency (e.g. Cook 2016; Jarvis 2015). Klein (1995, 426) maintains that if the second language was learned in a school-like setting instead of the home, the benefits are greater. It is unclear, however, how much knowledge in another language is needed for the experience to be useful in further language learning. Sometimes the L1 has less influence on learning a third language (L3) than the L2 does, especially if the L2 and the L3 are typologically similar; however, no matter how dissimilar the L1 and the target language are, learners tend to assume semantic and pragmatic similarities between the two (Ringbom and Jarvis 2009, 108; see also Benati and Angelovska 2016; Jarvis 2015; Klein 1995, 423-429; Ortega 2009; Ringbom 2016, 42-43; Rothman, Cabrelli Amaro and de Bot 2013). Learners' reliance on perceived similarities enables transfer, the application of prior knowledge to a new situation (see Section 2.3.2). However, learners differ in how readily they are able to make associations between their L1 and L2, and while learners of related languages can perceive similarities, learners of unrelated languages can only assume them (Ringbom 2016, 40-43).

In second language learning, the similarities between the first and the second language can be divided into actual similarities and assumed similarities, and learners can take advantage of either. According to Ringbom and Jarvis (2009, 107; see also Jarvis and Pavlenko 2010, 176-182), assumed similarities relatively rarely are congruous with actual similarities; moreover, learners' perceptions of assumed similarities change as they become more proficient. Similarity is actually a continuum with three types: similarity, contrast, and zero relations (see Figure 3).

Full-scale similarity is rare, "except in closely related languages that are mutually comprehensible" (Ringbom and Jarvis 2009, 109). Contrasts are typically linguistic features that learners readily perceive as differing in an important way. Here, learners are "aware of the existence of a system" (Ringbom and Jarvis 2009, 110), and once learners have discovered how a feature is expressed in the target language, they can relatively easily understand how to apply this knowledge. Zero relations are more challenging, as learners have difficulty perceiving any relation to the languages they know. This is the case with learning unrelated languages at the early stages. Thus, starting to learn an unrelated language requires more effort than learning a related language, because the learner has to struggle to understand how the new language works (Ringbom and Jarvis 2009, 110; see also Herschensohn 2013, 332-334). Later, when the learner has more knowledge, zero relations may become contrasts.

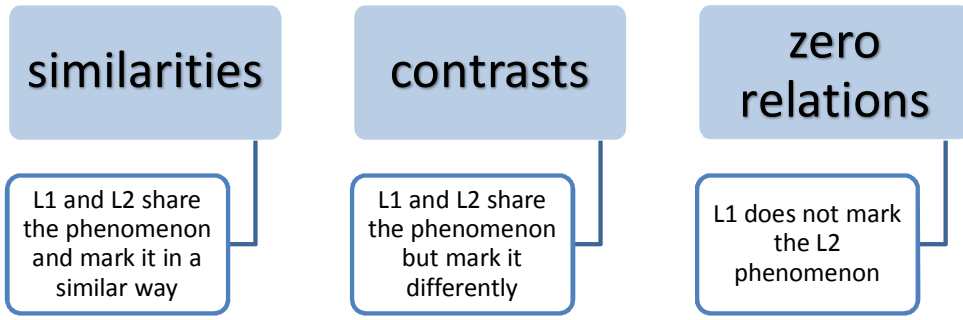


Figure 3. The relationship between grammatical phenomena in an L1 and L2

To investigate to what extent knowing other languages besides the L1 assists in the learning of subsequent ones, Klein (1995) studied how acquisition results in English differed depending on whether English became the learner's second or third language. She found that most multilingual adolescents outperformed unilingual students both in their acquisition of lexical items and in their syntactic learning (Klein 1995, 437-439), but some multilingual learners did not seem to benefit from their previous language learning experiences. Nonetheless, Klein (1995, 451) argues that the multilinguals were not really better at learning the new language: they were just faster at learning, which does not guarantee differences in ultimate attainment (for a discussion, see Lardiere 2013). However, Jarvis (2015, 73) maintains that any level of bi- or multilingualism can be helpful in the acquisition of additional languages because of enhanced cognitive abilities. This is typically manifested in tendencies, patterns and preferences, not in foolproof systems (Jarvis 2015, 74).

2.3.2 Factors that affect second language learning

Language acquisition is rarely linear: learners may be using both standard and unconventional forms in parallel, and with time, the proportion of standard usage begins to prevail (e.g. Ellis and Barkhuizen 2005; Jarvis and Pavlenko 2010; see also Thewissen (2013) in Section 4.4.2). As explained in Section 2.3.1, when learning a new language, people naturally rely on the information they have of other languages they know, their L1 in particular. The reliance on previous information in a new situation is called transfer (e.g. H. D. Brown 2007a, 102; for criticism of the term, see Cook 2016). In language studies, transfer is also called cross-linguistic influence and refers to "the learner's reliance on conceptions of both formal similarities across individual items and functional equivalences between the underlying systems" (Ringbom and Jarvis 2009, 112). There can be positive transfer, where previously learned information helps in learning the new task, or negative, where previous knowledge is not beneficial for performing in the new task but perhaps counter-productive, resulting in errors (H. D. Brown 2007a, 102; see also Alanen 1997, 57; Saville-Troike 2012, 19). Such negative transfer is called interference (e.g. Yule 2014, 191; for criticism of the term, see Ortega 2009, 31). While transfer can be both facilitating and interfering, teachers and testers tend to focus on negative transfer much more strongly, because it is manifested in errors (H. D. Brown 2007b, 76). The terms are summarised in Table 1, and errors are discussed further in Section 3.2. For an extensive discussion of crosslinguistic influence, see Jarvis and Pavlenko (2010) and Odlin (2003).

Table 1. Concepts related to the influence of the L1[6] in learning the L2

Concept	**Meaning**
Transfer	The learner relies on the L1 (or previously learned features of the L2) to learn the L2
Interference	The negative effects of reliance on the L1
Overgeneralisation	The learner uses L2 features in situations where they do not apply
Avoidance	The learner does not use an L2 feature

In addition to transfer and interference, second language learning can be characterised by overgeneralisation and avoidance. Overgeneralisation refers to extending the use of some features to contexts where they do not apply (e.g. Benati and Angelovska 2016, 89; Ortega 2009; Yule 2014, 176). This happens when learners apply a rule they have acquired in a

[6] While the influence of L2 on L3 and any Ln has been acknowledged (see Section 2.3.1), I only refer to L1 and L2 in this table to avoid making it overly complex.

situation where it is inappropriate because the learners do not (yet) know that there is something different in the new situation that prevents the use of this prior knowledge. An important related phenomenon is avoidance: learners might not use a grammatical structure, for example, because they are not confident with it, because the structure does not exist in their first language or because they perceive the structure as redundant (H. D. Brown 2007a, 259; Ellis and Barkhuizen 2005, 70; Gass and Selinker 2001, 119-120; Ortega 2009; Ringbom 2016, 39; see also Larsen-Freeman 2009, 521). Avoidance easily results in underuse of the difficult form and overuse of another form (Ortega 2009, 41-42).

Second language learning is not straightforward. Some structures may appear simple but be difficult to learn, and some complex structures may actually turn out to be easy (Nassaji and Fotos 2011, 136). Similarly, some structures that are different between the first and the second language may be easy to learn, and some structures which are similar may be difficult to learn (Hedge 2000, 170). Figure 4 illustrates some of the factors that influence whether L2 grammar topics are difficult to learn. For example, difficulties can arise from the meaning of the grammatical element, its form or the interplay between the form and the meaning (DeKeyser 2005). Whether the form, meaning or both are complex affects how well and how fast learners acquire them. When the relationship is transparent or salient, it is easier for the learner to acquire (DeKeyser 2005, 14; see also Jarvis and Pavlenko 2010; Salaberry 2000). A structure can also be difficult because the meaning of the structure is novel, abstract or both (DeKeyser 2005, 5). For example, it can be difficult to learn the aspect system in English if the learner's first language does not mark verbs with aspect. This is discussed further in Section 4.1, and the specific difficulties L1 Finnish and L1 Swedish may pose on learning English is discussed in Section 5.1. The ways how morphological features influence acquisition order is further discussed in Section 4.4.

Becoming native-like in an L2 is not easy, and many studies argue that very few learners' ultimate attainment results in native-like skills (see e.g. Herschensohn 2013, 321-330; Lardiere 2013). Although post-adolescent L2 learners "typically do not achieve nativelike convergence in all respects for the target language grammar they are acquiring" (Lardiere 2013, 671), some do. However, even such successful learners' performance may display variation that is not common in native speakers. Ultimate attainment, in this understanding, is the "state of knowledge actually attained at a stabilized endpoint of development in a particular domain" (Lardiere 2013, 670). Such endpoint may contain both successful and erroneous features of the target language, including fossilised items (Myles 2013; see also Han 2011; Long 2003 and

Section 3.2.1). For problems with the concept of ultimate attainment, see Whong and Wright (2013, 76).

Figure 4. Factors that affect how difficult it is to acquire grammatical structures in an L2

2.4 Summary

Variation is evident in people's use of a particular language, both by native and non-native speakers. In addition to regional variation, there is historical, socioeconomic, cultural, individual and register variation. Variation is inevitable, and the nature and range of variation can cause emotive response in some people. While some native speakers may resist cooperation with non-native speakers because of their inadequate language skills or because of prejudice, the responsibility for successful communication rests on both parties.

Learning a new language is challenging. Successful acquisition is influenced by many factors, as are the specific difficulties in learning the new language. Language learners may take advantage of their first language and other languages they have already acquired. While some features in this process are beneficial, some linguistic features do not tally, which can cause

problems in communication if the learners cannot distinguish between similar and dissimilar structures in the two languages. When the structures are fairly similar, this results in positive transfer, while negative transfer is manifested in errors. Learners may also avoid using structures they are not confident with.

In the case of English, the fact that there are several widespread varieties makes it difficult to decide which variety should be taught to learners or who qualifies as a teacher or as a native speaker of English. Today, with non-native speakers outnumbering native speakers, the ability to communicate successfully has become more significant than trying to meet prescriptive norms. This implies ever-increasing variation but still presupposes mutual intelligibility in order for successful communication to take place. Furthermore, this suggests that teachers are to reconsider which aspects of language they should focus on. For English as a lingua franca purposes, the focus may be on communication and increased fluency, but for academic purposes, more focus on accuracy may be important as well.

3 Grammar and errors

The role of grammar has long been one of the most controversial topics in language teaching (e.g. Nassaji and Fotos 2011, 1; Thornbury 1999, ix), at least partly because people understand grammar in different ways. The narrow understanding of grammar as simply syntax and perhaps also morphology is challenged by views that include the text level, discourse level and meaning-making potential of language (Thornbury 1999, 1-13), which may imply different pedagogical strategies. Consequently, the scope and precise nature of teaching grammar depends on the teaching and learning context (for an overview, see e.g. Nassaji and Fotos 2011, 121-134).

A discussion on the role of grammar in language teaching is needed in order to understand the background of the students and teachers in my study (see Chapter 7), because the results depend on how teachers understand grammar and acceptability and how they apply their understanding in rating the students' responses. There is a wealth of literature on grammar and its place in language teaching, and, for this reason, only a small section can be referred to in this chapter. For extensive overviews of studies on the role of grammar in teaching L2 English, see e.g. Ellis (2006), Larsen-Freeman (2009) and Lightbown and Spada (2006).

People understand the concept of grammar in different ways (e.g. Keck and Kim 2014; Larsen-Freeman 2009, 518). For instance, grammar can be understood as referring to the order of elements in a sentence, to following standard conventions in a language or to being able to produce understandable language (e.g. Thornbury 1999). Grammar can also be understood as being prescriptive (or normative), which would be, for example, following rules that are taught at school, or as being descriptive, which refers to actual language use (Gass and Selinker 2001, 7). Watson (2015, 10) argues that the prescriptive image that some teachers have of grammar "may relate to the way in which grammar is discussed in public discourse, with an emphasis on rules and error correction".

Grammar can also be understood as being related to meaning, discourse or style (Hedge 2000, 152-158). For any aspect of grammar, learners need to acquire the structural meaning, how it is formed and how it is used (Celce-Murcia and Larsen-Freeman 1999, 4; Larsen-Freeman 2009, 521), as very few 'rules' are context-free (Celce-Murcia 2002, 121). In Alho and Korhonen's (2018) definition, grammar is a complex, dynamic web of forms and meanings, within which

people speak, write and mediate. Ultimately, grammar ensures mutual intelligibility, without which communication becomes difficult.

These different understandings of grammar reflect differences in attention to the various levels of language, i.e. single words, sentences or larger chunks of conversation or text, including the social and pragmatic contexts. After a traditional view of understanding grammar as the way forms are arranged and patterned, as the forms or structures that are possible in a language or as a synonym to well-formed language (Thornbury 1999, 1-3), a wider understanding of grammar has emerged, including attention to the pragmatic, functional and discourse-specific features of language. Figure 5 illustrates the different levels of focus with the term grammar.

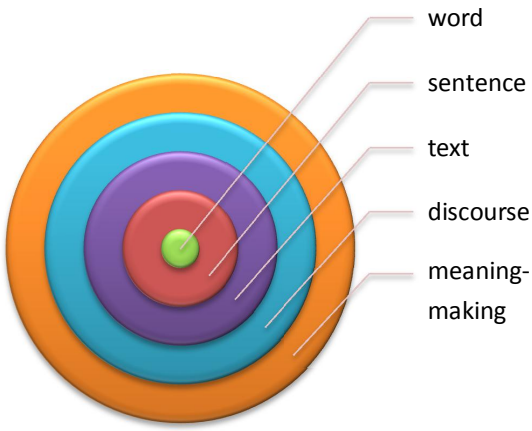


Figure 5. Different levels of understanding the term grammar

In the empirical part of this study, the approach to grammar that is used to inform my analysis of the students' responses (see Chapter 7) is derived from the teachers' ratings and therefore their understanding of acceptability in grammar (see Chapter 9). My own definition of grammar entails all of the levels described in Figure 5, but with the understanding that in specific contexts, the focus of grammar instruction may cater to particular aspects and levels depicted therein. I am particularly drawn to the idea of grammar offering choice (Larsen-Freeman 2002) and allowing the making of "decisions in an effective manner", as articulated by Celce-Murcia (2002, 121). In this study, however, the testing context limits what aspects of grammar can be explored (see Chapter 6).

In this chapter, Section 3.1 explores the influence of teacher cognition on the teaching of grammar. Section 3.2 focuses on various approaches to errors and discusses some challenges with testing students' skills in grammar, and Section 3.3 introduces the ways grammaticality and acceptability judgements have been used in research.

3.1 Teachers' views on grammar

Learners are usually not alone in the acquisition process: many second language learners study in a classroom setting and have a history of teachers accompanying them. In such cases, what teachers think about grammar matters, as this governs what they convey to learners; whether learners acquire or become aware of this is another matter. This section explores research on teacher cognition (Section 3.1.1) and discusses the impact of both positive and negative teacher attitudes on grammar instruction (Section 3.1.2). The beliefs that the teachers in my study expressed are discussed in Section 9.3. For a thorough discussion of research on teacher cognition, see Borg (2006; 2015), and for research on various ways to teach grammar, see Larsen-Freeman (2009).

3.1.1 Language teacher cognition and grammar

Research into language teacher cognition is a fairly new but intense area of study within applied linguistics. Borg (2015, 488-489) argues that teachers' own experiences as learners are powerful and may outweigh the effects of teacher education when teachers decide what to teach and how. These experiences tend to become established early and be persistent (Borg 2015, 488-489), although teachers can, of course, consciously try to direct their behaviour towards new paths. In the following, I limit the discussion to teacher cognition research in relation to grammar, which is one of the most researched areas of teacher cognition (Borg 2006, 109).

Summarising findings from several studies beginning in the 1980s, mostly from the United Kingdom, Borg (2006, 110-112) reports that many studies conducted on mainly pre-service teachers' declarative knowledge of grammar have shown significant gaps in knowledge, including misconceptions about language and inadequate metalinguistic skills. In some studies, non-native teachers scored much better than native prospective teachers. This may be because

non-native teachers have acquired normative rules to a greater extent than native speakers, who may rely exclusively on intuition. However, declarative knowledge is only "one component of the more global knowledge a language teacher must call on in teaching grammar" (Borg 2006, 112), and teaching grammar may not be the highest element on teachers' priority lists. For example, a study focusing on the amount of time devoted to various aspects of language teaching with in-service comprehensive[7] school teachers of English in Finland found that about 80% of teachers provide vocabulary instruction during every lesson and 20% every week, while about 40% of teachers provide instruction on grammar every day and 60% every week (Tuokko 2000, 111).

Borg (2006) also provides an overview of research on teachers' beliefs about grammar teaching, with conflicting results. A high number of teachers seem to be mainly influenced by their own language learning experiences (see also Barcelos and Kalaja 2013). In many studies, students requested error correction much more often than teachers were willing to provide it; furthermore, students were more interested in being taught grammar explicitly than teachers were willing to provide such instruction (Borg 2006, 114-115; see also Jean and Simard 2011). Thus, a mismatch seems to exist between student expectations and teacher beliefs (see also Section 3.1.2). Some other studies reported by Borg (2006, 115-116), however, showed that students did not master the terminology in grammar to the extent that teachers expected.

In addition to research on teachers' grammar-related beliefs and knowledge, there are studies on teachers' classroom practices, summarised by Borg (2006). Again, teachers seemed more influenced by their own experiences than by awareness of research on second language acquisition (Borg 2006, 120). However, although a teacher may decide to engage in explicit formal instruction, this does not necessarily imply that the teacher would believe that such instruction promotes language learning. Teachers may teach grammar because they find it important, because they think students expect it, or for reasons ranging "from any combination of acquisition-related, awareness-raising, diagnostic, psychological and classroom management factors" (Borg 2006, 124; see also Larsen-Freeman 2009).

Nevertheless, the impact of past experiences on teaching is not straightforward (Borg 2006, 124): teachers may teach grammar inductively, for example, exactly because they once learned the language through this approach and found it successful, or because they did not find another

[7] This is the term used by the Finnish National Agency for Education to refer to schools in compulsory education, grades 1-9. For more information, see the Finnish National Agency for Education and Chapter 5.

approach, for example the deductive approach, successful in their own learning process. Furthermore, teachers do not always teach in the way they self-report (Borg 2006, 126-127; see also Section 3.1.2). External reasons, for example institutional pressure or cultural expectations on how strictly curricula are to be followed, may make teachers behave in particular ways in classrooms (e.g. Barcelos and Kalaja 2013).

3.1.2 Positive and negative views on teaching grammar

There are differences in teachers' attitudes to teaching grammar, which, among other factors, stem from a different understanding of what grammar is. While the division into positive and negative views never affects an entire group of teachers or contexts, Figure 6 lists some generalisations that can be formed based on such views. While some studies in Western countries suggest that teachers are reluctant to teach grammar, studies in non-Western countries seem to indicate the opposite. In these cultures, teaching may be dominated by a focus on accuracy (Hyland and Anan 2006, 515; Thornbury 1999, 27; see also Shibata 2010 in Section 2.2) and reliance on elements from traditional grammar-based approaches (e.g. Fotos 2002, 142-143; Nassaji and Fotos 2011, 135), which can also influence and be influenced by learner expectations (e.g. Hinkel and Fotos 2002, 10; Thornbury 1999, 17). Such cultural differences may also be reflected in how teachers react to errors (see Section 3.2).

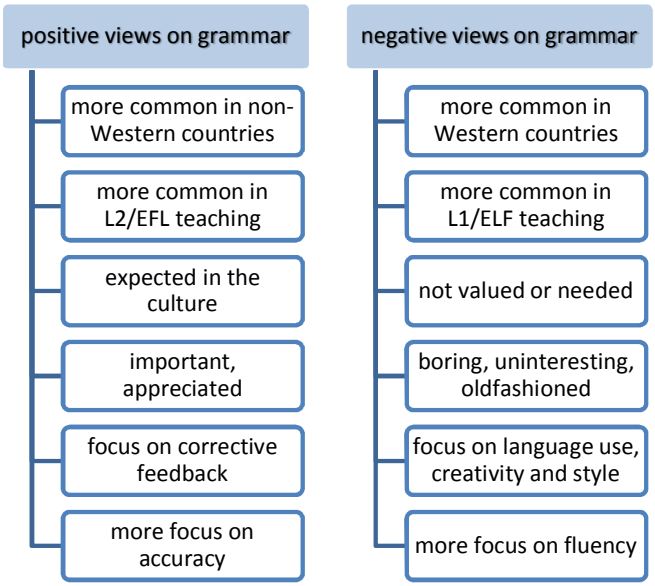


Figure 6. Generalisations related to positive and negative approaches to teaching grammar

For example, a study on non-native EFL teachers' grammar beliefs and classroom practices in Turkey shows that teachers seemed to value grammar teaching, and students were tested on grammar "to prove their language proficiency" (Hos and Kekec 2014, 83). In Turkey, a traditional approach to teaching grammar prevailed. However, while the teachers stated that they would not correct students' errors unless they were unable to convey the intended message, classroom observation proved otherwise, and although the majority of teachers claimed they taught according to communicative language teaching principles, classroom observation showed that the grammar-translation method with mechanical drills still prevailed; this was also the method through which most of the teachers had learned English. Thus, there was a mismatch between many teachers' beliefs and their actual practices (Hos and Kekec 2014, 84-85).

Similar results were found by Samad and Nurusus (2015), who explored teacher cognition in Malaysia with secondary school ESL teachers. According to the study, novice teachers did not pay enough attention to giving students feedback, while "more experienced teachers focus on corrective feedback in a classroom situation that emphasizes student learning through exposure to meaningful language input" (Samad and Nurusus 2015, 264). The researchers argue that greater focus on grammatical accuracy would be both necessary and expected in this culture.

In contrast to the views presented above, a number of studies in Europe and North America (e.g. Alho and Korhonen 2018; Jean and Simard 2011; Watson 2012; 2015) indicate that many teachers do not enjoy or see much value in teaching grammar, particularly when teaching native speakers (see Figure 6). This may be because teachers are unsure about what aspects of grammar to teach or how to do that (e.g. Watson 2012; 2015), possibly because they have not been taught grammar themselves, or because their own experiences of having been taught grammar were negative (Larsen-Freeman 2002, 103; Pennington 2002, 77-78). In addition, they may hold a narrow view of grammar and prefer teaching that focuses on communication without an explicit focus on grammar (e.g. Andrews 2007, 33; Nassaji and Fotos 2011, 6-9). For example, in Swan's (2002, 148-152) opinion, the only good reasons to teach grammar are to increase the comprehensibility and acceptability of the learners' output: if learners do not master the central structures of a language, they cannot truly communicate and may not be taken seriously by their interlocutors.

Teachers whose context mainly includes teaching native speakers seem to be particularly resistant to teaching grammar (e.g. Alho and Korhonen 2018; Watson 2012; 2015). Watson, who interviewed secondary school teachers of English in Britain, found that half of the teachers

described their feelings about grammar in negative terms, while many also "expressed some negative emotion, typically a lack of confidence in their subject or pedagogical knowledge which made them feel uneasy about teaching it" (Watson 2012, 29). In addition, many teachers considered grammar teaching boring and old-fashioned (see also Watson 2015, 6-9). Many of these teachers had not had explicit teaching in grammar when they were at school themselves. Only a few teachers found grammar interesting, even empowering, and many such teachers were self-taught in grammar (Watson 2012).

In a later study, Watson (2015, 1) further reports that many teachers only view grammar in a traditional prescriptive light and react negatively to teaching it. She believes that in order to make teachers more enthusiastic about teaching grammar, they need to be made aware of the variety of ways the concept of grammar can be understood. Many teachers in the study seemed unwilling to teach grammar because they perceived little value in it (Watson 2015, 6-9). When asked to define the term grammar, more teachers discussed the term with reference to rules, labels, accuracy or correctness than those who used terms such as rhetorics, effect or style. However, when the teachers were asked about the benefits of teaching grammar, they focused more on aspects of choice than accuracy or rules (Watson 2015). Thus, there is a discrepancy between what teachers say when asked about their beliefs and their perceived values. Watson (2015, 12) suggests that "it would be helpful to draw clear distinctions, for example, between the conventions of linguistic etiquette and the genuine patterns that underlie language, between descriptive and prescriptive grammar, between grammar taught to broaden the range of stylistic choices open to writers and grammar taught to improve accuracy in the use of standard written English". In Finland, Alho and Korhonen's (2018) findings indicate that primary school teachers understand grammar in three ways: 1) as a system of rules, 2) as a guide to spelling and 3) as command over a language. While Finnish teachers were not quite as negative about grammar as their British colleagues (Watson 2012; 2015), many teachers reported that they tended to avoid using the term 'grammar' in teaching but that they used the term when expicitly comparing linguistic phenomena in different languages (Alho and Korhonen 2018).

Partly similar results were obtained by Jean and Simard (2011), who studied high school students' and teachers' perceptions about grammar in second language education, with L2 French and L2 English in Canada. In their study, more ESL than French as a second language students found it important to be accurate, while the reverse was true for the teachers. ESL students were also more prone to demand corrective feedback than students of French (Jean and Simard 2011, 473-474). The authors speculate that this was because the ESL students were

accustomed to a rule-governed approach in their first language, French. Many of the students would have liked to have all their errors corrected, while teachers were more inclined to do so only when errors "impede comprehension" (see also Section 3.2.1) or when they concern "grammar points that should be known" (Jean and Simard 2011, 474; see also Borg 2006). Although many of these students and teachers acknowledged the need to discuss grammar in language teaching, they appeared to treat it as "a necessary evil" and did not enjoy it (Jean and Simard 2011, 478). The authors find it important to provide more successful pedagogical approaches to grammar, since "students' retention rate and motivation in L2 programs may be affected" by the negative attitude that they may notice in their teachers (Jean and Simard 2011, 467). Alho and Korhonen (2018) also believe that more guidance should be given on how to teach grammar.

Despite the negative views of grammar, recent research suggests that teaching grammar is beneficial for students' progress (e.g. Fotos 2002), even within a communicative approach, and particularly in the case of adult learners (e.g. Ellis 2002b). For example, Larsen-Freeman (2002, 104) argues that grammar "affords speakers of a particular language a great deal of flexibility in the ways they can express propositional, or notional, meaning and how they present themselves in the world", and that more focus should be given not only to the meaning and form but to the use of grammatical structures (Larsen-Freeman 2002, 116). Jarvis (2015, 83) further argues that such instruction would help learners enhance their "metalinguistic awareness, their knowledge of how various notions are conventionally expressed in the target language, and their desire to use the target language accurately and appropriately". For research on and ideas for teaching grammar, see e.g. Brown and Larson-Hall (2012), Celce-Murcia (2002), Celce-Murcia and Larsen-Freeman (1999), Fotos (2002), Keck and Kim (2014) and Larsen-Freeman (2002; 2009). Section 9.3 discusses the approach to grammar that the teacher participants in my study reported.

We need to remember, however, that teaching grammar does not equal learning grammar: "learners do not always acquire what they have been taught" (Ellis 2006, 86; see also Brown and Larson-Hall 2012). The shift from a prescriptive approach to a more descriptive one, combined with a shift in teaching towards a more communicative and interactive approach, may have changed the position of grammar in teaching as well as the understanding teachers have of grammar. Recent research supports "communicative focus on form" (Nassaji and Fotos 2011; see also Keck and Kim 2014) combined with a focus on meaning (Thornbury 1999, 25; see also Larsen-Freeman 2009, 525-526; Storch 2015, 353). This entails "instruction that draws

the learner's attention to linguistic forms in the context of meaningful communication" (Nassaji and Fotos 2011, 10; see also Brown and Larson-Hall 2012), awareness- or consciousness-raising in language instruction (e.g. Andrews 2007, 16; Bardovi-Harlig and Dörnyei 1998; Ellis 2002a; Thornbury 1999, 24) or 'grammaring', "the ability to use grammar structures accurately, meaningfully, and appropriately" (Larsen-Freeman 2009, 526).

3.2 Reactions to learner errors

Language learners may produce systematic errors in grammar, some of which can be explained by first-language influence. However, the order of acquiring grammar is not dictated by the order in which grammatical structures are presented to learners or how frequent they are (Schachter 1998, 558; see also Ortega 2009, 51-52). It can be difficult to identify the source of a particular error (Ellis 2002b, 27), and learners' ultimate understanding of concepts in a language may not match that of native speakers of the language. Furthermore, it can be very difficult to determine "what an error is an error of" (Gass and Selinker 2001, 82). Error analysis research, popular mainly in the 20th century, has shown that learners of dissimilar language backgrounds tend to make similar errors in learning a particular second language (e.g. H. D. Brown 2007a, 257; see also Ellis and Barkhuizen 2005; Saville-Troike 2012). This indicates that there are some universal tendencies in second language learning (see Section 4.4), but the learner's L1 strongly influences second language acquisition as well. One of the problems with applying error analysis is the fact that the data do not include the items that learners consciously or subconsciously avoid (e.g. H. D. Brown 2007a, 259; see also Sections 2.3.2 and 4.3); it is difficult to say much about items that are not used at all.

This section on errors is divided into two parts. Section 3.2.1 discusses the nature of errors and their origin from the learners' perspective, while Section 3.2.2 focuses on the position of errors in teaching and testing and examines inter-rater differences. For an extensive discussion on whether errors in grammar should be corrected in classroom settings, see e.g. Ellis et al. (2008).

3.2.1 Errors in learning

As we saw in Section 3.1, learners often expect error correction (or negative feedback, a term used by e.g. Ortega 2009), but some teachers are unwilling to provide it. Nonetheless, errors are both natural and inevitable in any learning process (Ellis 2002b, 22). H. D. Brown (2007a, 257-259) differentiates between mistakes and errors, arguing that mistakes are random lapses or slips made by both native speakers and learners. They do not show any lack of command per se; rather, they are momentary breakdowns in language production, and people can notice these themselves and correct them (Gass and Selinker 2001, 78-79). Errors, then, are the systematic display of "noticeable deviation from the adult grammar of a native speaker" (H. D. Brown 2007a, 258; see also Alanen 1997; Gass and Selinker 2001). They indicate that the learner has made assumptions about the second language that are unlike the practice of native speakers. However, it can be difficult to ascertain whether a particular unusual item is an error or a mistake, and often only a more systematic study of patterns is needed to distinguish between the two. Today, there is growing awareness that errors are "a clue to the active learning process being made by the student as he or she tries out ways of communicating in the new language" (Yule 2014, 191). The way errors are understood in my study is discussed in Chapter 7.

Some scholars distinguish between overt and covert errors. Overt errors take place at the sentence level and result in ungrammatical sentences, whereas covert errors are discourse level errors: they are grammatically correct but used in unsuitable contexts (H. D. Brown 2007a, 260; Ellis and Barkhuizen 2005, 56; cf. the difference between grammaticality and acceptability in Section 3.3). Furthermore, errors can be either global or local: global errors make communication difficult, while local errors only show a minor disturbance in communication. In addition, an error can be the addition of an unnecessary element or the omission of a necessary element (H. D. Brown 2007a, 262-263; see also Ellis and Barkhuizen 2005).

If learners fail to recognise their own errors and do not take advantage of potential feedback (cf. Ellis et al. 2008), this can result in fossilisation (e.g. Finegan 2004, 561; Long 2003; Thornbury 1999). The term refers to the persistence of an erroneous feature in perhaps otherwise fluent use of the second language. This is readily witnessed in the case of learner accents as well as persistent grammatical or lexical errors. Items become fossilised if students do not learn to correct their errors through the feedback they receive, either because they are not given constructive feedback or because they fail to pay attention to it (Finegan 2004, 561).

Fossilisation is never global. Instead, it is always limited to some linguistic phenomena, for example inflectional morphology, and the learner can excel at another aspect of language, for example syntax (Lardiere 2013, 685-691). Fortunately, however, fossilisation is not "some sort of terminal illness", and the learner can still make progress and ultimately acquire standard usage (H. D. Brown 2007a, 270; see also Ortega 2009). Thornbury (1999, 117) argues that some explicit attention to grammar ('focus on form') is necessary to ensure that fossilisation does not take place. For an overview of research on fossilisation, see Han (2011) and Long (2003).

3.2.2 Errors in teaching and testing

Teachers treat learners' errors in grammar in different ways (for an overview, see e.g. Ellis and Barkhuizen 2005, 173-175; Thornbury 1999). Teachers vary in their leniency as well: for some raters, any error is an error, while others create scales that treat some errors as graver than others. Furthermore, the same error may be evaluated very differently depending on the context. This variation in teacher approach to errors is one of the research questions in my study (see Chapter 6). For various approaches to treating errors in teaching, see e.g. Brown and Larson-Hall (2012), Ellis (2006), Keck and Kim (2014) and Ortega (2009); for the debate on the benefits of corrective feedback in instruction, see Ellis et al. (2008); for a meta-analysis on the effect of corrective feedback, see Russell and Spada (2006); and for an overview of the history of research on assessing learner knowledge, see Norris and Ortega (2011) and Saville-Troike (2012).

Some studies have found that native and non-native raters of learner language may react differently to non-standard phenomena, so that non-native raters focus more on adhering to norms, while native raters are more concerned with comprehension. For example, Hyland and Anan (2006) explored different raters' perceptions of error, using one Japanese EFL student's writing as the source and asking both native English teachers, Japanese EFL teachers and native English non-teachers to identify and correct the errors in the text and to give their reasons for the corrections. The results indicate that native English speakers were more lenient than Japanese teachers in grading the errors (Hyland and Anan 2006, 512). The Japanese teachers were more likely to employ the criterion 'infringement of rules' (e.g. inappropriate application of a rule), while native English non-teachers tended to use the criterion 'unintelligibility' (e.g. ambiguity or flow hindrance); English teachers employed both criteria.

In the study, Japanese teachers found agreement and word form errors the most serious, while native English speakers felt that word order errors were the most serious (Hyland and Anan 2006, 513). The Japanese teachers also found many more non-target errors (i.e. errors that the researchers had not anticipated) than the native English speakers, and the Japanese teachers were less consistent in their grading than the native English speakers (Hyland and Anan 2006, 517). The non-target errors included acceptability violations that can be categorised as having either a stylistic (levels of formality), discourse (cohesion and organisation) or semantic (lack of clarity) focus (Hyland and Anan 2006, 514). The researchers argue that the Japanese teachers' lack of exposure to different registers and their lack of confidence would explain why they take "a prescriptive attitude to correctness and a reluctance to accept non-standard forms" (Hyland and Anan 2006, 515). However, the researchers acknowledge that this may reflect cultural differences in teacher expectations and believe that teachers should be "more aware of the distinction between grammatical error and stylistic difference" and that more attempts at harmonisation need to be undertaken so that the same performance would not be assessed in varying ways (Hyland and Anan 2006, 518). While full agreement is impossible, rater training can minimise the effect of raters' overall tencency to be either lenient or strict (Kondo-Brown 2002, 4). For a discussion on harmonisation and standard-setting, see Fulcher (2016).

Although Hyland and Anan (2006) discuss the differences between Japanese and native English teachers' rating of errors, they only do so at the level of the number of errors spotted and do not discuss the nature of variation in the teachers' acceptance of such errors, and while they comment on inter-group differences in marking criteria, they do not discuss intra-group variation. In contrast, my study provides an insight into the variation that exists within teachers' assessment.

If different raters react to the severity of errors differently, so do learners and their teachers in different contexts. A study by Bardovi-Harlig and Dörnyei (1998) showed that EFL learners and teachers and ESL learners and teachers reacted differently to errors, depending on the nature of the error. In the EFL context, both learners and teachers found grammatical errors more serious than pragmatic errors, whereas the opposite pattern prevailed in the ESL context. Bardovi-Harlig and Dörnyei (1998, 247-252) argue that EFL learners and teachers focused more on structural accuracy, while ESL learners and teachers paid more attention to situation-specific appropriateness. However, both of the learner and teacher groups were successful at identifying errors; they simply rated their seriousness differently. Bardovi-Harlig and Dörnyei (1998) suggest that the difference may reflect ESL learners' greater access to relevant, authentic

input, as EFL learners may not have sufficient access to situations of authentic language use. Therefore, Bardovi-Harlig and Dörnyei (1998, 255) maintain that instruction should focus more on pragmatic awareness.

Many language tests focus on testing grammar (e.g. Thornbury 1999). What makes testing problematic is the fact that raters do not always behave according to expectations. Thus, "rating always contains a significant degree of chance" (McNamara 2000, 37). There are always borderline cases, and raters may disagree on whether a student's production passes the crucial threshold or not; raters are not always self-consistent, either. Indeed, according to McNamara (2000, 38), "[r]esearchers have sometimes been dismayed to learn that there is as much variation among raters as there is variation between candidates". Obviously, there are ways to alleviate the effects of variation, such as rater training and moderation meetings, where disagreement is discussed, consensus is sought and attention is given to how specific criteria are to be interpreted. There are also statistical tests that can be used to analyse inter-rater reliability (e.g. Salkind 2006; 2008). Nonetheless, even despite rigorous attempts at objectivity and harmonisation, intensive training and clear instructions, testers can still deviate from one another, for example in their level of leniency or their tendency to focus on particular aspects of the phenomenon being tested (e.g. Kondo-Brown 2002). This can also be influenced by the testing context, culture and goals. For a discussion on concepts related to harmonisation and standards-based assessment, see Fulcher (2016).

Raters' leniency or severity has been investigated by, for example, Huhta et al. (2014), who note that, inevitably, some raters are more lenient than others when rating students' performance, and some raters may be inconsistent in their assessment. For example, Huhta et al. (2014, 312-33) found that the difference in rater severity was half a scale point on a six-point scale, and 1.5-2.5 levels on a 10-point scale. The researchers also found that raters used scale-external criteria in their assessment when they felt that further criteria were needed. Some raters had more aberrant ratings than others and some raters were biased for some task types or particular rating scales, although no clear pattern was apparent and the rating seemed quite idiosyncratic (Huhta et al. 2014, 315). However, because removing a rater who significantly differed from the others did not change the overall ratings, "there appears to be no reason to consider removing raters simply because of their severity or leniency" (Huhta et al. 2014, 313). Furthermore, differences in ratings may also be a sign of difficulty in interpreting the scale. Huhta et al. (2014, 319; see also Kondo-Brown 2002) call for even more systematic benchmarking across countries and languages and highlight "the importance of having multiple

ratings of the learners' performances, as an individual rater's personal approach to rating a particular task (or scale) may bias the results". For this reason, my test includes thirteen raters. The extent of variation among the teachers in my study is discussed in Chapters 7, 8 and 9.

Another study on rater bias by Kondo-Brown (2002) found that raters' bias patterns were not uniform and that raters tended to treat some candidates and criteria more leniently than others despite the fact that they were often self-consistent and that their scores correlated with those of the other raters. The greatest difference in severity vs. leniency was found with the best-achieving and the low-achieving learners, and all raters had a unique bias pattern: one was harsher when rating vocabulary, while another judged errors with mechanics more harshly and the third rated content errors more harshly (Kondo-Brown 2002, 22). The raters discussed discrepancies after the first rating round, and the differences became slightly smaller in the following rounds, which means that harmonisation attempts slightly decreased the gap between the lenient and strict raters (Kondo-Brown 2002, 18). However, the study only included three raters, which is too small a number for any generalisations, and the texts the students wrote were often very short, sometimes only one sentence. Nonetheless, Kondo-Brown (2002, 25) suggests that rater training would help raters become self-consistent, although it cannot fully eliminate inter-rater differences.

3.3 Grammaticality and acceptability judgements

This section discusses one way of exploring the extent of variability in languages: people's judgement of the grammaticality or acceptability of particular ways of using the language. Grammaticality and acceptability studies have been used to assess linguistic reasoning and learners' and speakers' command of a language. Such studies are often grounded on Universal Grammar, which is a highly controversial theoretical framework. While my study does not contribute to this tradition and while it does not stem from Universal Grammar, it is necessary to explain what grammaticality and acceptability judgement studies are like so that it is easier to understand the difference. Universal Grammar itself is not discussed here; for extensive reviews, see Benati and Angelovska (2016), Gass and Selinker (2001), Ionin (2012), Klein (1995) and Mackey and Gass (2005). This section presents some results gained from grammaticality and acceptability judgement studies, both from studies on the first language

(below) as well as studies on second language learners' judgements of acceptability (Section 3.3.1). Acceptability judgement studies have also faced fierce criticism, which is discussed in Section 3.3.2.

Some scholars (e.g. Alanen 1997; Bader and Häussler 2010; Dąbrowska 2010) distinguish between grammaticality and acceptability: grammaticality refers to a sentence conforming to the rules of grammar, while acceptability refers to a sentence being permissible in the language (Dąbrowska 2010, 4; cf. the difference between covert and overt errors in Section 3.2.1). Thus, a sentence that is grammatical may not be acceptable, for example, because it violates the semantic conventions of specific words, or because it is overly complex. Ultimately, a sentence does not really have any value if it is grammatical but not acceptable (Bader and Häussler 2010, 277). In addition, a sentence can be unacceptable for other reasons, for example because it burdens the memory too much or because it is pragmatically implausible (Dąbrowska 2010, 4; cf. Bader and Häussler 2010). This implies that several factors affect where the borders of acceptability lie, including the pragmatic context, the semantic content of the sentence, the grammaticality of the sentence, the judge's knowledge of the world and the discourse setting. Some scholars have investigated whether a difference can be found between acceptability and grammaticality judgements, and for example Moreno et al. (2010, 574) found that bilinguals were less accurate than monolinguals in acceptability tasks, but reached a comparable level of accuracy in grammaticality tasks. Although my study uses the concept of 'acceptable' (see Chapter 6), it studies the phenomenon from a very different point of view (see Chapter 1).

It seems that grammaticality judgement tests are best interpreted as a continuum, not as a yes/no decision. Bader and Häussler (2010, 317) found that all of the testing methods they used led to similar results: participants gave similar judgements independent of whether there was a time pressure or not and whether they were only provided yes/no alternatives or an extensive scale to judge the sentences. Bader and Häussler (2010, 321) acknowledge the fact that people "sometimes are unsure about whether a particular sentence is grammatically correct or not". There seeems to be some correlation between acceptability and frequency in use so that if structure A is less frequent than structure B, it is never more acceptable than structure B, but the reverse is not necessarily true (Bader and Häussler 2010, 315-316).

3.3.1 Grammaticality and acceptability judgements in second language studies

Grammaticality and acceptability judgement studies have also been conducted with L2 learners to assess whether they "have the same abstract representations as do native speakers", by forcing "learners into stating what is possible and what is not possible in their second language" (Mackey and Gass 2005, 49; see also Ionin 2012). Dąbrowska notes (2010, 5) that "acceptability judgments are routinely used in the L2 literature, and are assumed to reflect L2 linguistic knowledge". This section presents some such studies. However, there has been debate on what grammaticality judgement tests in second language studies actually show (e.g. Ellis and Barkhuizen 2005) and whether they are a valid and reliable tool to describe learners' language skills; this is discussed in Section 3.3.2.

Some grammaticality and acceptability studies compare learners to native speakers. For example, Robenalt and Goldberg (2016, 74) explored whether native and non-native participants differ in the acceptability judgements of novel sentences and argue that while native speakers readily take competing formulations into account, non-native speakers are not equally skilled in this. They also found that learners whose self-reported spoken skills in English were the highest were also the closest to behaving in a native-like manner in the experiment (Robenalt and Goldberg 2016, 80). Alanen (1997) found that native speakers were both faster and more accurate in their judgement than learners, but some of them also made errors. She also found that the most accurate learner judgements were given by learners who "had been exposed to English primarily in formal language learning settings" (Alanen 1997, 162), and more proficient students were faster in their judgements. In addition, learners were faster at judging grammatical than ungrammatical sentences.

Grammaticality judgements have also been used to investigate the influence of age and first language on the acquisition of an L2, for example to explore whether there is a critical period within which native-like mastery in an L2 can be acquired (McDonald 2000; for the critical period in language learning, see e.g. Lightbown and Spada 2006, 67-74 and Herschensohn 2013; for an extensive critical overview, see Singleton and Muñoz 2011; for ultimate attainment in an L2, see Lardiere 2013). In McDonald's study (2000, 404), Spanish-speaking students of English who had moved to the United States and started English before the age of 5 scored very highly, while the scores of the students who had moved after the age of 14 varied significantly.

The Vietnamese-speaking students in McDonald's (2000, 410) study did not score as well as the Spanish-speaking ones, although also here the early acquirers outperformed the late acquirers. McDonald (2000, 413) concludes that the Vietnamese learners' problems were the most prominent in the areas where the Vietnamese and the English language differ (cf. Section 2.3 and Chapter 4). Similar results were found by Alanen (1997): Finnish students were the least accurate with ungrammatical sentences when they concerned the errors that are typically made by Finnish learners of English.

As mentioned above, the validity of grammaticality judgement tests has been questioned (see Section 3.3.2). However, Mandell (1999) defends the view that they are a valid approach to assess students' skills, supporting this with evidence from tests on English native speakers' scores in L2 Spanish. He asked both beginner, intermediate and advanced students to complete grammaticality judgement tests, where students indicated whether specific sentences were possible or impossible in Spanish, and dehydrated sentence tests, where students were asked to create acceptable sentences from a given set of words. His findings indicate that the two tests yield similar scores across skill levels, which, in Mandell's (1999, 93) opinion, implies that grammaticality judgements are a reliable measure of students' command of an L2. However, such a setting seems to test learners' command of standard word order rather than their overall skills. For other merits with grammaticality and acceptability judgement tests, see Whong and Wright (2013).

Some other researchers have found that grammaticality judgements need to be applied with great caution. For example, Davies and Kaplan (1998, 196-197) found that native and non-native speakers use different strategies when deciding on the grammaticality of particular sentences. While the most common strategy in judging grammaticality in the L1 was intuition, in the L2 context it was more common to apply a rule the participants had learned or simply to guess. Meanwhile, Tabatabaei and Dehghani's (2012) results indicate that students were not consistent in their judgements over time, changing about 35% of their initial responses when re-tested later (see also Alanen 1997, 71). The researchers report that when participants were uncertain about the grammaticality of some sentences, they were unwilling to label them as 'not sure' but preferred guessing (Tabatabaei and Dehghani 2012, 181). Both Davies and Kaplan (1998, 198-199) and Tabatabaei and Dehghani (2012) conclude that grammaticality tests should be used with extreme caution, as they do not seem to be reliable in identifying learners' linguistic ability. I return to this at the end of Section 3.3.2.

3.3.2 Criticism on grammaticality and acceptability judgements

As we have seen, grammaticality and acceptability judgement studies have been used to investigate phenomena in both native speakers' and learners' language use. However, a number of problems have been associated with such studies (e.g. Alanen 1997; Dąbrowska 2010; Ellis and Barkhuizen 2005, 19-21; Gibson and Federenko 2013; Myles 2013, 55-56; Whong and Wright 2013). A summary of such criticism is provided in Table 2.

First of all, the overall validity of Universal Grammar is questionable: according to Dinsmore's (2006, 80) meta-analysis, Universal Grammar "does not fully operate in adult/adolescent L2 learning". Furthermore, it now seems that there are very few universals across languages (e.g. Cook 2016, 25). Another criticism is that grammaticality judgements have often included measuring reaction times of participants' identification of grammatical or ungrammatical sentences (e.g. Ellis and Barkhuizen 2005). Quite often, participants have had to make their judgement under time pressure and with isolated sentences, without any context. The practical value of such studies is questionable. Furthermore, learners participating in grammaticality judgement tests are "influenced by non-linguistic factors, such as learners' varying attention or willingness" to make judgements (Norris and Ortega 2011, 578).

Table 2. Criticism of grammaticality and acceptability judgements

Nature of criticism	**Reason for criticism**	**Corrective measures**
Isolated, decontextualised sentences	Specific contexts may require different solutions	Provide more context
Too few judges	Enables bias and unjustified generalisation	Use more judges
Experts used as judges	Experts have repeated exposure to linguistic forms that are rare in actual use	Use naïve judges
Yes-no scale	Acceptability is a gradient phenomenon	Use a more extensive scale
Artificially created sentences	Little value in studying contrived sentences	Use more natural sentences
Unreliable	Impossible to say what the judgement is based on; allows guessing	Exercise great caution in reporting the results

Another problem with grammaticality and acceptability judgement studies is that grammar is not rigid. There are instances where "otherwise unacceptable utterances become acceptable in a given context" (Boas 2011, 1272). In addition to the conventionalised senses of particular items, it is sometimes possible to find unconventional usages in specific situations and contexts (Boas 2011, 1275-1284). Furthermore, if the judgement is simply between grammaticality and ungrammaticality, it remains difficult to prove what the learner's criterion is for the judgement (Dinsmore 2006, 59).

A further problem arises from the number of judges used to testify to the appropriateness of particular expressions. Many studies assessing learners' ability of judging the grammaticality or acceptability of sentence/meaning pairs have used either very few or even single judges, often only the writer of the article (Gibson and Federenko 2013, 89; see also Dąbrowska 2010; cf. Section 3.2.2). However, if this one judge is given unlimited power, it may lead to unjustified conclusions. When the use of one judge is combined with a small number of participants and a small number of stimuli, the cognitive biases on the part of the researchers can become overly prominent and prone to overgeneralisations and faulty interpretations (Gibson and Federenko 2013, 89).

Gibson and Federenko (2013, 98) also argue that since expert informants (e.g. experienced linguists) typically have cognitive biases related to the judgement of the acceptability or grammaticality of certain items, it would be better to use naïve participants. In particular, expert informants are "biased due to their understanding of the theoretical hypotheses", i.e. they may deduce which phenomenon is being investigated and respond according to what they think the researcher expects (Gibson and Federenko 2013, 99; see also Dąbrowska 2010). Furthermore, such informants may have greater exposure to particular types of structures and potentially prescriptive attitudes to the stimuli, while naïve informants are more likely to respond "based on their own intuitions and are not affected by cues from the experimenter" (Gibson and Federenko 2013, 101). According to Gibson and Federenko (2013, 116), there are several cases where "questionable intuitive judgement has led to an incorrect generalisation, which has then led to unsupported theorising that many researchers have followed".

While the above studies focus on research in testing situations, research on authentic student production and on second language learning often also uses single informants, typically the students' teacher or the researcher. For example, Tremblay (2011, 351; see Section 6.4.2) does this: she is the researcher, teacher and the only judge in determining the acceptability of learner

French. Similarly, many studies that assess learners' skills in a foreign language rely on error-tagging, with predefined decisions on what constitutes an error; an example of such practice is the study by Murakami and Alexopoulou (2016), cited in Section 4.4.1. In such cases, limitations in the teachers' or researchers' knowledge and their language-related biases may distort the results and assessment.

Repeated exposure may also affect people's judgement so that if they see a structure, even an inappropriate one, reappear, they may start finding it more acceptable. For example, Dąbrowska (2010), who studied naïve and expert judges' intuitions in grammaticality or acceptability judgements, found that linguists' judgements were black-and-white, while the naïve informants displayed a continuum. Dąbrowska (2010, 15) argues that "linguists' judgments are sensitive to grammatical structure and relatively insensitive to lexical content, while the opposite is the case for the nonlinguists". She further argues that linguists may have a different view on certain structures compared to naïve informants simply because they have seen more such structures; some structures that are prevalent in literature on linguistics are fairly rare in actual use (Dąbrowska 2010, 15-21). Hence, experts "cannot simply rely on their own intuitions and assume that they are representative of the community at large" (Dąbrowska 2010, 21). To some extent, teachers may have a similar bias.

As was discussed in Section 3.3.1, great caution needs to be exercised when applying the results of grammaticality and acceptability judgement studies. A related problem is that it remains uncertain what the studies try to tap (e.g. Alanen 1997; Ellis and Barkhuizen 2005; Mackey and Gass 2005, 50). If a learner marks a sentence as unacceptable but continues to produce similar sentences, what does this tell us about the learner's understanding of the second language? Furthermore, what is the learners' criterion for the judgement: idealised use, actual use or something else? Thus, using acceptability judgements remains somewhat problematic because very little can be said about the learners' skills based on how they rate isolated sentences, focusing on the appropriateness of sentences that may have been artificially created (for a discussion, see Mackey and Gass 2005).

As can be seen, grammaticality and acceptability judgements usually expect learners to work under time pressure and judge the appropriateness of isolated, artificially created sentences without any context. In contrast, my study (see Chapter 6) is very different in that it provides a coherent text with contextual cues and exerts no time pressure on the students; they simply fill in the gaps to the best of their ability. Furthermore, it is not the students who are to assess the

grammaticality or acceptability but the teachers, and they work within a specific context, taking as much time as they need. Nevertheless, I acknowledge the influence of the criticism presented above in some features of my study. For example, I do not claim that my test evaluates the students' overall proficiency in English, and I use 13 teacher informants to avoid bias from a limited number of raters. Furthermore, the teachers are provided with a gradient scale instead of a black-and-white assessment criterion. In addition, the test does not consist of isolated sentence pairs but provides a contextualised story.

3.4 Summary

Grammar is understood in many ways, from a narrow view focusing on syntax to a view focusing on the meaning-making potential of language. Studies on teachers' beliefs and their attitudes to grammar seem to produce conflicting results. It may be that teachers whose own background contained detailed attention to grammar continue to provide such instruction to their students, while teachers whose attention was not drawn to grammar may feel uncomfortable with providing grammar instruction. Non-native teachers are perhaps more prepared to teach grammar as their background more likely included extensive instruction in grammar, while native teachers may not have experienced much grammar instruction at all. Furthermore, teachers in non-Western countries are perhaps expected to teach grammar explicitly, while many teachers in Europe may be less willing to do so. In addition, there can be some mismatch between teachers' and learners' understanding of the term 'grammar'.

While teachers may not enjoy teaching grammar or correcting errors, learners often tend to expect instruction in grammar and explicit error correction. Moreover, if learners' attention is not drawn to any systematic errors they may make, some of these features may become fossilised. While fossilisation is not necessarily permanent, recent research suggests that learners would benefit from explicit attention to grammar.

In testing students' skills in grammar, different raters do not always agree on the rating to be given to a particular response. Despite attempts at standardisation and extensive training, raters may continue to assess students' production in different ways. Ultimately, even the best raters have occasional 'blind spots', when they either lose concentration or occasionally take a more severe or lenient approach compared to their general performance as a rater.

Grammaticality and acceptability judgements have been used to assess students' proficiency in grammar, but the approach has some problems. The results cannot be easily generalised, and they may not truly reflect what the participants can do in a language. Moreover, if there is only one judge determining whether the sentences are grammatical or acceptable, there is a danger of bias, which may lead to ill-justified conclusions about students' ability or acceptable language use.

4 Verbs in English

Since my study focuses on Finnish students' use of English verbs, this chapter discusses previous research on the acquisition and use of verb forms in English. Because this comprises such a large body of research, only a selection of relevant studies can be introduced. The features of verb use discussed below are chosen in light of what the test I use requires (see Section 6.1). However, many of the studies referred to here address lower levels of proficiency, and there are fewer studies on the use of verb forms at the university level. For an extensive discussion on verb forms in English, see Biber et al. 1999.

In this chapter, I discuss studies on the acquisition of tense and aspect (Section 4.1), the spread of the progressive (Section 4.2) and verb errors in academic writing (Section 4.3). In Section 4.4, I present some theories on the order of acquisition in learning English verb forms and an attempt to place the order on an assessment scale. The verb forms the students in my study provided are discussed in Chapters 7 and 8.

4.1 The acquisition of tense and aspect

Many studies have been conducted on second language learners' acquisition of tense and aspect in English. The most prominent approaches are the aspect hypothesis (for a thorough overview, see e.g. Gass and Selinker 2001; Keck and Kim 2014; Muñoz and Gilabert 2011) and the discourse hypothesis (e.g. Bardovi-Harlig 1998). Both of these hypotheses are introduced below and illustrated in Figure 7. The success of the students in my study in providing tense and aspect is discussed in Section 8.1, but there is no attempt to test for the validity of either hypothesis[8].

The aspect hypothesis claims that "learners make initial associations between tense-aspect markers and prototypical semantic categories for their use" (L. Collins 2004, 253). When students progress in their studies, they learn to expand their knowledge to less prototypical

[8] Due to limitations of scope, I do not address issues related to the aspect and discourse hypotheses in the discussion of results.

contexts. According to the aspect hypothesis, learners typically use the simple past[9] to code events, the progressive to discuss activities and the simple present to mark states (Bardovi-Harlig 1998, 475).

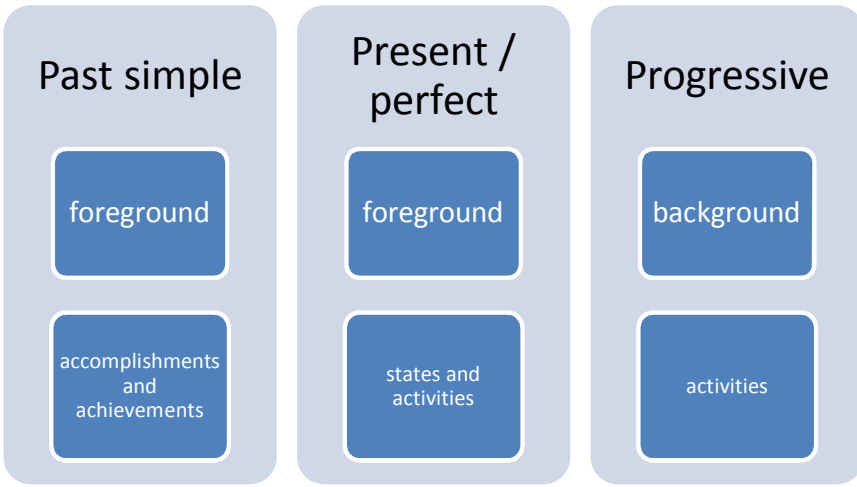


Figure 7. A summary of the aspect and discourse hypotheses

According to the aspect hypothesis, the first verb types that are marked for past tense are telics, i.e. verbs that indicate an inherent end point (e.g. *to finish*). Such verbs are sometimes divided into achievements and accomplishments. Telics are followed by atelic verbs (process verbs with no inherent end point, e.g. *to read*), while stative verbs (e.g. *to love*) are the last to be marked for past tense (Salaberry 2000, 137). Gass and Selinker (2001, 156) summarise the findings of the aspect hypothesis as follows:

1. Past / perfective morphology emerges with punctual verbs and verbs indicating achievements and accomplishments. The morphology then gradually extends to verbs expressing activities and states.
2. Imperfective morphology emerges with durative and / or stative verbs (i.e., activities and states), then gradually spreads to achievement / accomplishment and punctual verbs.
3. Progressive morphology is strongly associated with durative and dynamic verbs (i.e., activities).

Similarly, studies on the discourse hypothesis have established that learners tend to use the simple form in the foreground in their narratives. If learners use progressive forms at all, they

[9] In this study, I use the terms 'simple past' and 'past simple' interchangeably.

are much more likely to appear in the background (Bardovi-Harlig 1998, 475-477). It seems that the two hypotheses are best combined to explain how learners acquire verbal morphology. According to Bardovi-Harlig (1998, 497-498), students rely both on lexical aspect (grouping verbs into states, activities and telics), which follows the aspect hypothesis, and on grounding (dividing narrative discourse into the foreground and the background), which follows the discourse hypothesis. However, acquisition is also influenced by the learners' first language and the way it codes tense, aspect and mood, and whether it does this at all. Because languages code linguistic phenomena in different ways, what is coded in verb forms in one language may be coded in another part of speech in a different language.

The hypotheses have been tested in several studies. For example, Muñoz and Gilabert (2011), in their study of bilingual Catalan-Spanish learners of English, note that the suffix *-ing* seems to emerge before the past tense suffix *–ed* (see also Section 4.4.1) and that the use of the progressive increased with the proficiency level (Muñoz and Gilabert 2011, 253-259). L. Collins (2004), who studied French- and Japanese-speaking learners of English to investigate whether the aspect hypothesis would hold with two very different first languages, found that both French- and Japanese-speaking students used the progressive forms more with activities than other semantic verb categories, and both language groups had similar rates of success with using the simple past (L. Collins 2004, 259-264). However, Roberts and Liszka (2013, 427) suggest that learners whose L1 has grammaticalised aspect are more sensitive to a mismatch in tense/aspect in English than learners with a L1 that does not code aspect in its verb forms. In their study of French and German advanced learners of English, Roberts and Liszka (2013, 431) conclude that the French learners were more sensitive to violations in the tense/aspect system because aspect is a grammaticalised element in French and in English but not in German.

Another study by L. Collins (2005) provides information on how retrospective reports can contribute to the understanding of the development of tense and aspect in learner English. In this study, participants were asked to articulate their hypotheses on how the English tense and aspect system works in order to determine whether the division of verbs into semantic categories such as telics, activities and statives has validity in learners' decisions on which tense or aspect to choose (L. Collins 2005, 210). Her findings with Japanese and Chinese learners indicate that students mainly rely on their knowledge of the prototypical uses of verbs. She argues that learners may be "exposed to a disproportionate number of perfective or simple past markers with telics, progressive with activities, and present with statives in native speaker input" (L. Collins 2005, 216; see also Muñoz and Gilabert 2011). Although prototypical

associations may facilitate learners in acquiring the tense-aspect system, such prototypes may also constrain the learners so that they are unable to acquire less prototypical cases.

4.2 The spread of the progressive

Although learners mainly rely on prototypical uses (see Section 4.1), there are innovative trends in the use of tense and aspect as well. The use of the progressive seems to be increasing (e.g. P. Collins 2008; Gut and Fuchs 2013; Leech et al. 2009; Rautionaho 2014; cf. Section 9.2.2), both within native varieties and in learner English. For an extensive overview, see Leech et al. (2009).

According to Gut and Fuchs (2013, 245), the "prototypical meaning of the progressive involves an activity without implicit or explicit endpoint that is expressed by a dynamic, durative, atelic verb". However, the progressive can be used in a variety of other meanings, including, for example, a reference to either limited time or a series of repetitions (for an extensive overview, see Gut and Fuchs 2013, 245). Verbs which take human subjects as the agent and describe prolonged situations are common with the progressive, while verbs that have human subjects as the experiencer and describe non-continuous processes tend not to be used in the progressive (Gut and Fuchs 2013, 249; see also Biber et al. 1999).

Overall, the frequency of progressive forms has increased in the past two hundred years (Leech et al. 2009). Studies on the spread of the progressive in English show that, in some cases, the Outer Circle varieties are more likely to create and accept non-prototypical uses of the progressive, such as those with stative verbs (P. Collins 2008; Gut and Fuchs 2013, 254; Rautionaho 2014). Nevertheless, there is more variation "across registers than across dialects" (P. Collins 2008, 247): for example, progressive forms are twice as frequent in speech as in writing, "more common in news than in academic prose, and [...] more common in fiction than in news" (P. Collins 2008, 248; see also Biber et al. 1999, 460-475). Moreover, the present progressive is used more in speech than in writing, but the past progressive is more common in writing than in speech. Similarly, Rautionaho (2014, 213-214) found that, of the eight varieties of spoken English she studied, the progressive is used most extensively in Irish English and least extensively in Hong Kong English. Furthermore, the Outer Circle varieties use the progressive to refer to the future more often than Inner Circle varieties (P. Collins 2008, 248;

Rautionaho 2014, 215). In general, the use of the progressive in its conventional functions is more common in British and American English than in other varieties (Rautionaho 2014, 215).

The spread of the progressive is not only limited to the Inner and Outer Circle varieties, but it seems to be happening in the Expanding Circle as well. For example, Edwards (2014) examined the various uses of the progressive in Dutch English and argues that Dutch English seems to display features of both second language and foreign language varieties: progressive marking showed "no evidence of over-reliance on the action-in-progress prototype" (Edwards 2014, 188). She notes that some forms that are considered 'innovative' in ESL varieties "tend to coincide with those held up as common 'errors' in EFL" (Edwards 2014, 174). Given the extent of innovations in these varieties, perhaps a new understanding of variation in grammar is needed to guide EFL instruction. The students' use of the progressive in this study is discussed in Section 8.1.2.

4.3 Verb errors in L2 academic writing

In addition to tests and research situations, students may have difficulty finding the appropriate tense and aspect in actual use contexts, such as academic writing. Students typically convey some writing conventions from their L1 to their L2 writing (e.g. Jarvis and Pavlenko 2010, 102-106), and since the students in my study (see Section 6.2) all attended university, they can be expected to have some command over academic writing (cf. Section 2.2). A pioneer in this field of research, Hinkel (2002; 2004; 2011) has studied the use of verb forms in academic texts created by non-native writers of English. She reports that L2 writers tend to use tenses inconsistently and to use fewer passive constructions than native English writers (Hinkel 2002, 182). Moreover, L2 writers' texts may lack subject-verb agreement, have incorrect or lacking morphology, employ modal verbs incorrectly and contain spelling errors (Hinkel 2011). None of this is surprising, however, given that they are writing in a second or foreign language.

Rather pessimistically, Hinkel (2011, 530) notes that even at the university level, "the micro properties of L2 writers' text continue to differ significantly from that of novice L1 writers in regard to a broad range of features" and that "even advanced and highly educated L2 writers [...] have a severely limited lexical and syntactic repertoire". Although such errors also occur in L1 writing, Hinkel (2011, 531) claims that they are "fundamentally distinct from those in L2 university writing because [L1 writers' errors] are unlikely to impede comprehension". Thus,

the author's expectations of what non-native writers should be able to do seem to greatly differ from what they can actually do.

In another study, Hinkel (2004) compared the usage of tense, aspect and the passive voice in university-level texts written by various Asian and Arabic[10] learners of English to such usage by native speakers of English. Her results show that most non-native speaker groups used past tenses significantly more frequently than native speakers and provided texts that were more like narratives than argumentative essays. Native English students tended to use the present tense in their texts and to rely on generalisations and observations, while non-native students often described their personal experiences (Hinkel 2004, 14-16), which indicates that the conventions associated with suitable academic style vary from culture to culture (Hinkel 2004, 17). She further observes that few non-native speakers used any progressive or perfect verb forms at all; the same is true for passive forms. This may be because the students wanted to "avoid using syntactically and semantically complex verb structures" (Hinkel 2004, 22). However, Crossley and McNamara (2011, 280) argue that since complex writing burdens the working memory more than choosing simpler structures, learners may choose to apply less demanding tactics in writing than L1 writers, who may have been able to automatise some of these processes. This would explain why L2 writers' texts are often less complex and syntactically and lexically more limited than L1 writers' texts (Crossley and McNamara 2011, 280).

Decisions on whether non-native students would benefit from further instruction in grammar depend on the instructional context. Meriläinen (2010b, 61) argues that a focus on communicative competence is not sufficient for strengthening students' skills in English grammar. If students are to be communicative, and if they are to be classified as lingua franca speakers (see Section 2.2), gaps in their mastery of English syntax may not be a problem. However, if students are to pursue academic studies and to write professionally, it would be in their interests to pay more attention to grammar in order to be more credible as academics (Meriläinen 2010b, 63). H. D. Brown (2007b, 422; see also Storch 2015, 353) also notes that more attention to grammar would be needed particularly in advanced and formal writing, such as academic writing, and Hinkel (2004, 24) believes that textbooks should stress the importance of the passive in academic writing (for passive forms in this study, see Section 8.3.3). She suggests that non-native writers be advised to "avoid using the future tense and instead use modal verbs" (Hinkel 2004, 25; for modal verbs in my study, see Section 8.3.1). Furthermore,

[10] Note that, typologically, these languages are very different from both Finnish and Swedish.

students should be advised not to use progressives in academic writing as they are rare in native use and many verbs cannot be used in the progressive at all (Hinkel 2004, 26; see also Sections 4.1 and 4.2). However, non-native writers' texts can easily become clumsy if they heavily rely on using formulaic language.

4.4 The development of accuracy

Although learning to use verb forms in L2 English follows a slightly different process depending on the learners' first language, scholars have suggested that an order of acquisition exists in learning at least some grammatical rules (for a discussion, see e.g. Ellis and Barkhuizen 2005; Gass and Selinker 2001; Keck and Kim 2014; Lightbown and Spada 2006; Ortega 2009; Ur 2011; Vainikka and Young-Scholten 2013). This means that certain structures are acquired in a particular order and that learners cannot acquire more complex or more advanced structures before the previous structures have been successfully acquired (Benati and Angelovska 2016; Ellis 2002a, 170; Gass and Selinker 2001, 317-320; cf. Pienemann 2005). At least to some extent, this order seems predictable, unmodifiable and independent of the learner's first language (Hedge 2000, 150). Although the students in this study are expected to have completed their path along these acquisitional ladders, it is important to know what these ladders are and to see if the students have been successful in their development of accuracy. In this section, I focus on developmental orders proposed for the acquisition of verbs (for other aspects of English, see e.g. Gass and Selinker 2001; Lightbown and Spada 2006). After a general discussion in Section 4.4.1, I focus on an extensive study by Thewissen (2013), who places the hierarchy on a widely used assessment scale, the CEFR, in Section 4.4.2.

4.4.1 The development of accuracy in verb forms

Many scholars (e.g. Cook 2001, 26-29; Finegan 2004; Gass and Selinker 2001; Yule 2014) argue that there is a hierarchy in how children learn some features of their L1 English and that this hierarchy also exists in L2 English, at least partly. However, although the order seems to be the same for all children, the speed at which the acquisition takes place varies (Finegan 2004, 548). In second language learning, Cook (2001, 27) argues that the same principle holds irrespective of the learners' first language (see also Myles 2013; Ortega 2009; for an overview

on research in the field, see Goldschneider and DeKeyser 2001). Other scholars argue that the L1 strongly influences the hierarchy, even to such an extent that no universal hierarchy exists (e.g. Murakami and Alexopoulou 2016; Luk and Shirai 2009).

Research suggests that the progressive *–ing* ending is easier for L2 learners to acquire than forms of the word *be*, and that with *be*, the copula appears before auxiliary uses (Cook 2001, 27-28; Hawkins and Casillas 2008; Young-Scholten 2013). All of these appear before the irregular past tense, which occurs before the regular past tense, which occurs before the third person *–s*; see Figure 8. The difficulty with acquiring the third person *–s* may be caused by the fact that it "conflates person, number, tense, and aspect" (Goldschneider and DeKeyser 2001, 36). Hawkins and Casillas (2008, 596-598) also argue that the use of *be* + bare verb (e.g. '*I'm read*'), which does not appear in native English, is a universal feature of learner English, independent of the L1. However, while there seems to be a 'natural' order of difficulty in learning morphemes in English, there is no agreement on why this might be so or whether it is a universal order (DeKeyser 2005; Gass and Selinker 2001). While Goldschneider and DeKeyser (2001) argue that a small number of determinants can explain an important part of the acquisition order, Murakami and Alexopoulou (2016) as well as Luk and Shirai (2009) argue that there is no universal order.

Figure 8. The assumed order of acquisition in verb forms (adapted from Cook 2001, 27)

Goldschneider and DeKeyser (2001, 1) argue that there are five contributing components that influence whether a grammatical morpheme is difficult to acquire: perceptual salience, semantic complexity, morphophonological regularity, syntactic category and frequency (see also Section 2.3.2). They maintain that the cumulative effect of these determinants explain a significant portion of the total variance in students' production (Goldschneider and DeKeyser 2001, 33-37). Nonetheless, they acknowledge that the five determinants can all be understood as aspects of salience so that the phenomenon would operate at various levels and facilitate acquisition. However, Goldschneider and DeKeyser's (2001) research focused solely on ESL studies and excluded studies conducted in EFL contexts, which may be a factor potentially compromising the generalizability of the results. Furthermore, they did not test for the effect of L1 transfer (see Section 2.3.2).

Some research, however, stresses the fact that the L1 strongly influences the order of acquisition (e.g. Murakami and Alexopoulou 2016; Luk and Shirai 2009). Such findings challenge the view that there would be a universal order for the acquisition of morphemes in English as an L2. The results gained by Murakami and Alexopoulou (2016, 394) showed that there was a "clear between-L1 difference in the accuracy order" with respect to all of the morphemes they studied, which were the same morphemes that Goldschneider and DeKeyser (2001) had used. Murakami and Alexopoulou (2016) showed that learners with different L1s learned English morphemes in a different accuracy order, and the result depended on whether comparable morphemes were present or absent in the learners' L1 as well as whether the present morpheme was optional or obligatory. They found that Spanish learners followed the pattern assumed to be the universal order, but Japanese, Korean, Russian, Turkish, German and French learners deviated from the assumed order of acquisition (Murakami and Alexopoulou 2016, 386; see also Luk and Shirai 2009). In particular, the acquisition of the progressive *–ing* was shown to be dependent on whether progressive aspect was marked in the learners' L1.

When referring to past events, Lightbown and Spada (2006, 91-92) argue that after a period of not marking time in any way (see also Ionin 2013, 509; Myles 2013, 52), learners begin to attach morphemes to verbs, but not necessarily the one that is used by native speakers for marking past time. For example, *-ing* may be used for marking past time instead of the standard *–ed*. Irregular past tense may appear before the regular past forms, but once the regular forms appear, their use may be extended to irregular verbs as well (Lightbown and Spada 2006, 91-92). However, languages may have different temporal distinctions, and the present perfect in English often creates particular challenges to learners (Celce-Murcia and Larsen-Freeman 1999, 124-125; Jarvis and Pavlenko 2010, 139-142). For research on students' difficulty with inflectional morphology in L2 English, see Lardiere (2013).

Some researchers (e.g. Hawkins 2007, 472-476; Salaberry 2000) argue that difficulty producing the regular past tense in English may stem from phonological constraints, for example where learners have an L1 that does not contain word-final consonant clusters. This may explain why some students more readily produce correct irregular forms compared to regular past tense forms. The same phenomenon may also explain why the third person singular *–s* is dropped by some learners (Hawkins 2007, 472-476; see also Ionin 2013): learners may ignore *–ed* or *–s* because they do not perceive them (Hawkins and Casillas 2008) or do not find them important. Furthermore, Ambridge (2010, 1499) argues that it may be easier for learners to acquire irregular verbs, because most of the commonly used verbs in English are irregular. When

learners acquire more verbs, the regular patterns become increasingly typical, and most verbs will ultimately fall under the regular category (Ambridge 2010, 1497-1499). For the success rate of the students in this study in providing irregular verb forms, see Section 8.3.2, and for overall success with the past simple, see Section 8.1.1.

Learners' order of acquisition may not be the order in which grammatical elements are presented in textbooks or in the classroom. Even if they are not easy to acquire, certain structures need to be introduced early in order to maintain meaningful conversation (Hedge 2000, 172). A classic example is questions: although their formation is complex, it is difficult to maintain a discussion or to teach without asking questions from early on.

4.4.2 The development of accuracy and the CEFR

One of the most detailed reports on developmental patterns in learner English has been conducted by Thewissen (2013), who studied the development of accuracy in EFL learners using a learner corpus. The corpus included native speakers of French, German and Spanish, whose skills in English ranged from B1 to C2 on the CEFR[11], the Common European Framework of Reference for Languages (CEFR 2001). Using the corpus, she was able to explore where progress[12] can be seen as learners become more proficient, which elements remain stabilised and where the critical points of development lie on the CEFR scale. Her results show three main developmental patterns: a) the strong developmental pattern, where there is "a statistically significant difference in behavior between at least one pair of adjacent proficiency levels"; b) the weak developmental pattern, where there is a marked difference "between at least one pair of non-adjacent levels" and c) the non-progressive trend, where error types show "no significant behavioral change from B1 to C2" (Thewissen 2013, 83-84). In the following, I only discuss the error types that involve verb forms. They are listed in Table 3.

Thewissen's study (2013, 85-88) indicates that within the strong developmental pattern (adjacent proficiency levels), the number of spelling errors, missing words, morphological

[11] The CEFR is widely used these days to assess students' language skills on a holistic framework. Although it was originally intended for self-assessment, it has been applied for other purposes as well. For a discussion on the role of the CEFR in Europe, including some criticism, see Cook 2011. For aligning tests to the CEFR, see Papageorgiou (2016).

[12] The CEFR itself avoids discussing grammar and focuses on communicative competence and on what learners can do in a language. This is because the CEFR is to work with any language, and it is impossible to provide language-independent criteria on the complexity of grammatical structures.

errors (leading to non-existent words) and errors in phraseological units decrease significantly when learners move from B1 to B2. In addition, this is the stage in proficiency development when verb number (verb agreement) errors, verb morphology errors and verb-dependent preposition errors markedly decrease. Although there does not seem to be statistically significant quantitative progress in these error types at levels B2 to C2, Thewissen (2013, 89) argues that there is qualitative progress, as students start taking greater risks and have command over increasingly complex patterns. Furthermore, in some cases significant learning has already taken place at the A levels, and the mean error percentage score may be minimal already at B1 (Thewissen 2013, 87).

Table 3. Developmental patterns in the acquisition of English verb forms based on Thewissen (2013)

Developmental pattern	**Feature**	**Relevant progress in CEFR levels**
The strong developmental pattern	spelling errors	B1 to B2
	missing words	B1 to B2
	morphological errors leading to a non-existent word	B1 to B2
	errors in phraseological units	B1 to B2
	verb number (verb agreement) errors	B1 to B2
	verb morphology errors	B1 to B2
	verb-dependent preposition errors	B1 to B2
	adverb order errors	B to C
The weak developmental pattern	word order	B1 to C1/C2
	verb voice	B1 to C1
	modal auxiliaries	B1 to C2
The non-progressive trend	tenses	-
	verb complementation	-
	finite vs. non-finite verb forms	-

In Thewissen's study (2013, 89), only one error type shows improvement from the B macro-level to the C macro-level: adverb order errors. This means that significant increase in proficiency appears only when looking at the entire B vs. C level. Furthermore, C-level learners typically use significantly more adverbs than B-level learners, which seems to point to "simultaneous co-development between accuracy and complexity" (Thewissen 2013, 89) as students begin to use more complex forms and are more accurate in their use of previously

learned forms. Thewissen (2013, 90-91) also identifies some weak developmental patterns, with statistically significant development between non-adjacent proficiency levels. This includes word order, with improvement mainly from B1 to C1 and to C2, and verb voice, with increased mastery from B1 to C1. Interestingly, modal auxiliaries show significant improvement only from B1 to C2. This may be because modal auxiliaries have complex uses, often affecting larger elements than just their immediate surroundings in a sentence (Thewissen 2013, 91).

Finally, Thewissen's (2013, 92) study includes error types that show no statistically significant improvement at all from B1 all the way to C2, despite some improvement. This group includes, for example, tenses, verb complementation and finite vs. non-finite verb forms. She notes that tenses remain a difficult part of English grammar, perhaps because students need to have command over both tense and aspect, and because tense choices often have implications over larger elements than single clauses (Thewissen 2013, 93). Furthermore, she thinks these error types may resist change because attempts at increased complexity do not show in the quantitative data. Thus, although students try to use increasingly advanced and complex forms and, hence, show signs of learning more, they may make more errors in using the more complex forms while becoming better at using the simpler forms they learned earlier, which results in similar scores in the number of errors despite improvements in quality (see also Storch 2015, 350). However, the total number of errors were small in both verb complementation and finite vs. non-finite verb forms, with mean error scores already below 0.5% at level B1 (Thewissen 2013, 92-93).

Based on the findings, Thewissen (2013, 94) finds further support for the argument that language learning is not a linear process (see Section 2.3.2): as a matter of fact, few error types showed only positive development. Several error types persisted with some progress and stabilisation, or with stabilisation only. The most marked development in accuracy took place between B1 and B2, which leads Thewissen (2013, 95) to suggest that this is a "possible accuracy threshold", after which accuracy remains fairly stable. Nonetheless, while progress slows down between B2 and C2, it does not disappear (Thewissen 2013, 96). However, the study only included learners at levels B and C and does not discuss the development that takes place already before learners reach level B1. Furthermore, the study only included German, Spanish and French learners of English and does not account for potential differences with learners from other L1 backgrounds. While I do not attempt to place the students in this study on the CEFR scale, the expectation is that they would have reached level B2 (see Sections 5.2 and 5.3). For students' command of verb forms in this study, see Chapters 7 and 8.

4.5 Summary

The English language poses challenges to L2 learners with its complex system of verb forms. In particular, the existence of regular and irregular verb inflection and the difference between simple and progressive forms may be difficult for learners. In addition to morphology, learners need to understand the appropriate contexts for particular forms, which is challenging even for advanced students. Genre-specific requirements, such as those in academic writing, pose further challenges. In particular, learners may struggle with tense, aspect and the passive. Further instruction would make L2 writers more credible as academics and help them follow the conventions of academic writing.

Research shows that there may be a hierarchy in the order of acquisition of verb forms, which is partly universal and partly dependent on the L1. For example, progressive forms take longer to stabilise than simple forms, and they may remain limited to the prototypical uses for a long time. Nevertheless, the increasing uses of the progressive imply that a greater range of uses is likely in both L1 and L2 English in the future. Generally, learners tend to associate verb types with particular forms, for example the progressive with activities and the simple with telics. Furthermore, the progressive is more common when providing the background in narratives, while the simple is more common in the foreground. The learner's L1 also influences which forms are difficult to acquire.

In English, a number of verb forms and uses can take a long time for learners to master. In some cases, significant improvement only appears at the C1 or C2 levels on the CEFR grid. However, the learning process is not linear, and learners may make quantitatively more mistakes when they produce qualitatively more complex structures. The major threshold in accuracy seems to be between B1 and B2.

5 Finnish learners and the English language

In the past few decades, Finns have earned a reputation for a high level of English skills (e.g. European Commission 2012; EF 2015). In particular, highly educated people, including Finnish university students, are assumed to have good skills in English. This chapter provides a historical, pedagogical and linguistic background to the participants in this study.

Since this study focuses on verb forms, Section 5.1 first briefly explores the main differences in verb forms between English, Finnish and Swedish. Both Finnish and Swedish are discussed because Finland is a bilingual country. Section 5.2 focuses on how the Finnish education system supports foreign language studies, after which Section 5.3 explains how national curricula describe the instruction that is to be provided in English lessons in upper secondary school, as this governs what skills students are expected to have by the time they enter university. Section 5.4 discusses Finns' skills and use of English in national and international comparisons, including findings from both holistic and more specific studies on Finnish students' English skills.

5.1 Verbs in Finnish, Swedish and English

This section provides a short summary of the main differences between English, Finnish and Swedish in how verb forms are formed and used. Only features relevant to verb formation and verb use are discussed here; furthermore, the discussion is limited to aspects of the construction and use of verb phrases that are different in the three languages, particularly from the perspective of the forms that are needed to respond to the test in this study (see Section 6.1). For a detailed discussion on the differences between English and Finnish, with examples, see e.g. Meriläinen (2010a). Section 5.1.1 compares verb formation and use in Finnish and English, while Section 5.1.2 compares Swedish and English.

5.1.1 English vs. Finnish verbs

There are significant differences in the use of verbs between Finnish and English. Finnish is an agglutinating and inflectional language (e.g. Finegan 2004, 60), which relies on inflectional morphemes, mainly suffixes, in its morphology; these suffixes are distinct and can readily be segmented into parts, and the inflectional system is elaborate (e.g. Ojutkangas et al. 2014). The details provided below are from an extensive grammar of Finnish by Hakulinen et al. (2004). The use of the symbol § below follows the policy that the book[13] uses. For a Finnish grammar in English, see Karlsson (2008).

In Finnish, verbs carry tense, mood and person, and they can occur in finite and non-finite forms, including participles (§ 105). The finite forms can be complemented with a variety of structures, but prepositions are rare (§ 687). Some verbs occur in idioms or are complemented with particles. Not all verb forms agree with a subject: there are also verbs that only occur in the third person singular and take no subject (§ 457). In negation, the negative word *ei* takes some of the inflections that verbs carry in affirmative sentences; some other inflections remain in the verb (§ 108; see also Karlsson 2008, 108). The passive verb form does not agree with a noun as the subject but uses an impersonal form (§ 110). The underlying principles of the formation and use of the passive are completely different in the two languages, but the concepts exist in both languages (see also Karlsson 2008, 249).

There are two morphological tenses, the present and past, and the combined tenses, the present perfect and the past perfect, are formed with an auxiliary (a form of the verb *olla*) preceding the main verb, similar to English (§ 450). However, the use of the past simple and the present perfect are not identical in the two languages. There are four moods: the indicative, conditional, potential and imperative, some of which are expressed using a modal auxiliary in English, while in Finnish they are formed with suffixes (§ 111). In the passive, a suffix but no auxiliary is needed for the present or the past, but in the perfect forms, the auxiliary *olla* is needed (§ 110). Furthermore, if the conditional is given with a perfect reference, the auxiliary *olla* is used (§ 112). There are no irregular inflections.

[13] Obviously, some phenomena are discussed in various sections in Hakulinen et al. (2004). However, as it would be impractical to refer to each of the sections that mention a particular phenomenon, I only provide the section that serves as the introduction or is the first of the sections devoted to the phenomenon.

A potential difficulty for Finnish learners of English arises from the fact that depending on the context and other words in the sentence, the word *olla* can either be translated into English as *be* or *have* both as the main verb and as the auxiliary. Some of the inflections are placed on the auxiliary. In Finnish, no progressive aspect appears in a grammaticalised form that would be directly comparable with English, but there are ways[14] to express that an action is (or was) ongoing (§ 1498, § 1519). The future is not grammaticalised, but there are various lexical ways to express future reference. Generally, however, the present tense is used with an adverb marking future time (§ 1542; see also Ojutkangas et al. 2014).

In Finnish, word order is more flexible than in English because the case system in nouns is used to indicate, for example, who does what to whom (§ 1221), and because the target and the active partner are expressed in cases, it does not always matter which is placed first. Differences in information structure are often encoded in word order (§ 871). Adverbs can also move relatively freely. In reported speech, the tense of the reported content does not change systematically as it does in English. Reported speech can be marked with finite and non-finite constructions, and in some of these, tense is not visible in the verb form that is used (§ 1459).

Thus, the two languages differ considerably in how the ongoing nature of an action is expressed, how the passive is understood and constructed, how negative sentences are created, when an auxiliary is needed and which parts of the inflection are loaded on the auxiliary, how the future is expressed, how reported speech is expressed and how words can be placed in relation to one another in a sentence. All of these can potentially cause difficulties for Finns when learning and using English.

5.1.2 English vs. Swedish verbs

As explained in Section 5.1.1, all Finns are familiar with both Finnish and Swedish. While only some people consider themselves bilingual, all Finns learn and use both languages to some extent (see also Section 5.2). Swedish and English are both Germanic languages, and since Swedish is typologically much closer to English than Finnish, some structures in English are

[14] These include structures such as *olla tekemässä (§ 1519)* and *tehdessään*. However, although they refer to ongoing activity and resemble the progressive aspect in English, these are not considered grammaticalised in the same way as in English, and their use is significantly less common than the progressive in English. Further, the extent of their potential use is much more limited, and sometimes they refer to uses where the action remains an intention (comparable to *I was just about to*) and does not refer to an ongoing activity.

easier for L1 Swedish-speakers to acquire. This section focuses on the differences between English and Swedish verbs. The details provided below are from an extensive grammar of Swedish by Teleman, Hellberg and Andersson (1999). The book uses chapter numbers followed by the symbol § for sections, and I follow their policy[15] here. For a Swedish grammar in English, see Holmes and Hinchliffe (2013).

In Swedish, verbs carry tense and mood but not person (7 § 34-42 and 31 § 1-6), and they can occur in finite and non-finite forms, including participles (7 § 34-42). The finite forms can be complemented with a variety of structures, and phrasal verbs are common (16 § 1-25). There are several ways to form the passive, one of which is similar to the system in English (7 § 45-49 and 34 § 1-29; see also Holmes and Hinchliffe 2013, 320). There are two morphological tenses, the present and past, and the combined tenses, the present perfect and the past perfect, are used with an auxiliary (a form of the verb *ha*) preceding the main verb, similar to English (7 § 29 and 31 § 1-31). Unlike in Finnish, where the word *olla* is used for both *be* and *have*, in Swedish the two are separate words, *vara* and *ha*. In Swedish, no progressive aspect appears in a grammaticalised form comparable to English, but there are other ways[16] to express that an action is (or was) ongoing (5 § 9 and 33 § 6; see also Holmes and Hinchliffe 2013, 287-290). There are both grammaticalised and lexical ways to express future reference (7 § 29 and 31 § 28-31; see also Holmes and Hinchliffe 2013, 294).

In Swedish, word order is stricter than in Finnish and partly resembles the English usage. However, inversion is more common and there are more cases where the verb precedes the subject than in English (28 § 1-13 and 35 § 1-24; see also Jarvis 2015, 78; Jarvis and Pavlenko 2010, 204-205). Reported speech mainly follows the same pattern as it does in English (see Holmes and Hinchliffe 2013, 298). The verbs that are irregular in English are mainly irregular in Swedish as well (7 § 56-60), and negation is performed with a lexical word, *inte*, which is never bound to the verb nor inflected, but it has a place of its own in word order (see also Holmes and Hinchliffe 2013, 568).

Thus, English and Swedish have more structural similarity than English and Finnish, but there are also challenges for Swedish-speaking learners of English. These include the extent of

[15] Again, some phenomena are discussed in various sections, and I only provide the sections that serve as the main introduction or are the first set of sections devoted to the phenomenon.
[16] This includes e.g. the structure *hålla på att göra något*. The structure is limited in use and sometimes ambiguous in whether the action remains an intention or is an ongoing activity. This structure is partly similar to the corresponding ones in Finnish (see Footnote 14).

inversion, word order, how to refer to the future, the use of the progressive, the passive and the extent to which some verb structures are applied. However, the likelihood of benefiting from crosslinguistic transfer is high.

5.2 The role of English in Finnish schools

This section focuses on the role of languages in the Finnish education system in order to understand what the educational background of the student participants is like. First of all, Finland is a bilingual country: the two official languages are Finnish and Swedish, while a few other languages (Sami languages, for example) have some recognised status as well. About 89% of the population has Finnish as their first language, and about 5% of the population speaks Swedish as their first language (Ojutkangas et al. 2014; Statistics Finland 2016). A majority of the Swedish-speakers are, in effect, bilingual. English has no official status in Finland, but it is widely used and studied, and its importance is continually growing (see Section 5.4). The Finnish education system supports language studies: pupils are expected to study at least one foreign language in addition to both Finnish and Swedish, and many schools offer courses in more than one foreign language. The following account describes what most of my participants (see Section 6.2) had studied[17] when they were in comprehensive and upper secondary school. For a detailed account of the present Finnish education system, see the Finnish National Agency for Education and Leppänen et al. (2011).

In the 1990s, the Finnish primary school lasted for six years and lower secondary school for three years. Pupils usually began primary school (grades 1-6) at the age of seven and their first foreign language in grade three, at the age of nine. This was called the A1-language, and it was English for most pupils. Some schools offered an optional second foreign language called the A2-language in primary school, usually from grade five. Pupils taking the A2-language were expected to proceed faster than the A1 pupils of the same language so that they would reach the same level by the end of primary school. At the age of 13, pupils moved to lower secondary school for grades 7-9. There they continued to study the A-language for another three years and also started studying what was known as the B-language or "the other domestic language", i.e. Swedish for Finnish-speakers and Finnish for Swedish-speakers[18], if they had not started this

[17] Although I use the past tense to explain the system, many parts of the education system remain the same today. However, some of the terminology and some details have changed.
[18] Most Swedish-speakers began their Finnish studies in primary school, some already in the first year.

earlier. Furthermore, pupils had the option of starting yet another language in grade eight; this was an elective language, a C-language. Finishing lower secondary school marked the end of compulsory education, and the pupils had studied their A1-language for seven years. The typical[19] language curriculum offered in the Finnish school system in the 1990s is illustrated in Figure 9.

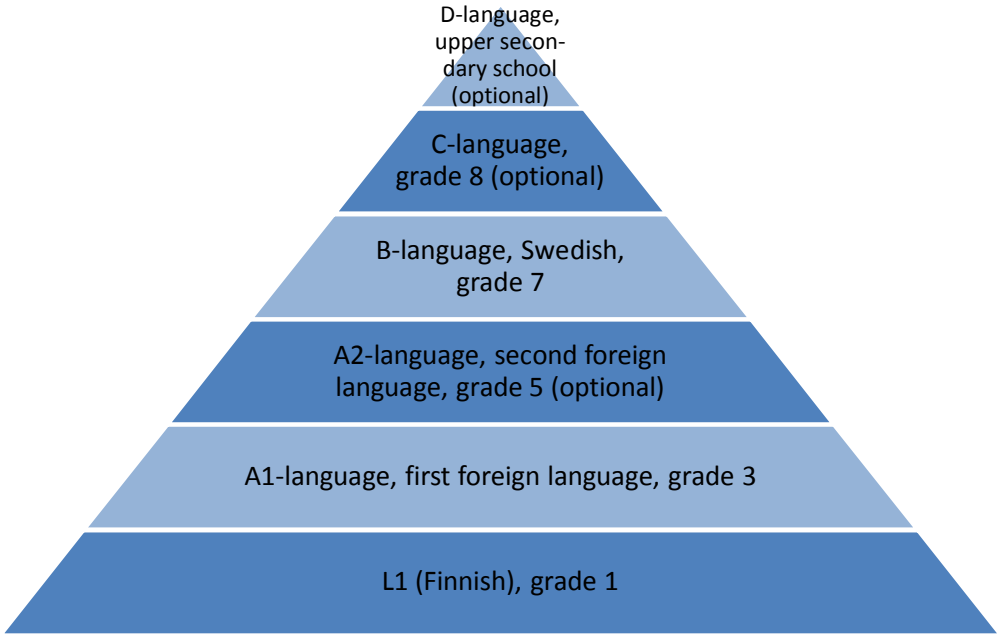


Figure 9. The standard starting grade of language studies in the Finnish school system in the 1990s from the L1 Finnish perspective

After comprehensive school, students with good marks often continued to upper secondary school. Upper secondary school typically lasted for three years, and offered the possibility of taking yet another foreign language, called the D-language, from the beginning of upper secondary school. Students usually finished upper secondary school at the age of 19. Thus, when beginning to study at university, the majority of the participants in my study (see Section 6.2) had studied English for ten years, while a minority had had eight years of English. Furthermore, they had all studied both Finnish and Swedish and possibly other foreign languages. However, several studies on L1 Finnish and L1 Swedish Finns' proficiency in English show that L1 Finnish-speakers tend not to reach the same level of skills in English as L1 Swedish-speakers do (e.g. Ringbom 2016, 49; see also Section 5.4.2).

[19] As mentioned above, some schools did not offer A2-languages, and many Swedish-speakers' language curriculum was a little different.

At the tertiary level, students were expected to complete studies in at least one foreign language in addition to both Finnish and Swedish as mandatory elements for their degree (see Section 6.2). A significant proportion of students completed the foreign language requirements in English, either by taking courses or by passing an exemption test. For details about the present system, see e.g. University of Helsinki Language Centre (2018).

5.3 The Upper Secondary School Curricula

In order to understand what the students in this study could be expected to master when entering the university, it is necessary to know what they were supposed to have been taught in upper secondary school. The goals and content for teaching English in upper secondary school is determined in *Lukion opetussuunnitelman perusteet* (Framework Curriculum for the Senior[20] Secondary School). The curricula that are the most relevant for this study are those from 1985 and 1994 (Kouluhallitus 1985 and Opetushallitus 1994), as most of the participants in my study had entered upper secondary school when one of these curricula was applied. In the following, I only focus on how the curricula describe teaching either any foreign language or English in particular, and on how the curricula describe teaching the foreign language/English as an A-language (see Section 5.2).

The 1985 curriculum is very thorough in listing what students are to be taught (Kouluhallitus 1985). It quotes as its aim that students should be provided with sufficient ability to understand and use language and promoting the maintenance and further development of the skills they have acquired. The central goal is the development of communication skills in core situations of language use: understanding and becoming understood. Furthermore, the curriculum elaborates that students must gain command of the structures of the English language so that they can apply their skills in new contexts (Kouluhallitus 1985, 61). The 1985 curriculum also states that students are to understand more difficult language than what they can produce themselves. Such statements cannot be found in later curricula.

[20] This term was used in a translation of the curriculum, but the term 'upper secondary school' is more commonly used.

In the 1985 curriculum (Kouluhallitus 1985), teachers are guided in the choice of topics and vocabulary in teaching English. There is an extensive list of structures that students are expected to be able to understand and/or use by the end of upper secondary school. As the list of structures for students to master and/or understand is exhaustive, I only focus on verbs in the following list. Furthermore, I only list what the students were supposed to be able to produce; there are even more structures that they should be able to understand. Moreover, this list only applies to students with English as an A-language (see Section 5.2). Thus, according to the 1985 curriculum, students were supposed to master the use of the following English verb forms:

Tense

- the simple present, past, present perfect and past perfect
- the progressive forms of the tenses above
- the future form, the auxiliaries *will* and *shall* in the first person questions
- the structure *be going to*
- the simple present to express future time in subordinate sentences with a condition and time
- the imperative

(Modal) auxiliaries

- the auxiliaries *can*, *could*, *must*, *may*, *mustn't*, *should* and the structure *have to* in different tenses
- the auxiliary *should* (*ought to*) in commands and negations
- the auxiliaries *can* (*could*) and *must* and their corresponding equivalents in their main meaning and in different tenses
- the auxiliary *may* (*might*) and the structure *be allowed to* in different tenses
- the structures *have/get something done* and *used to*

Conditional sentences

- the first and second conditional
- the irreal conditional when referring to both the present, future and past

The passive

- the passive in different tenses, both with and without the agent
- the passive formation type where the dative becomes the subject of the passive sentence
- the passive with the auxiliaries *can*, *must*, *may*, *should* and *ought to*
- the verbs *suppose*, *say*, *think*, *expect* in the passive
- the words *we, you*, *one*, *they* and *people* in active structures with a passive meaning

The infinitive

- the main cases of the infinitive with and without the word *to*
- the most common infinitival structures
- the infinitive in shortened sentences
- some specific infinitival structures[21]
- the present infinitive with sensory verbs (*hear*, *sec* [sic!], *watch* etc.) and with *let* and *make* in both active and passive sentences
- the perfect infinitive in the active
- the negative infinitive in the most important cases

The gerund

- the gerund with the most common verbs requiring it (e.g. *avoid*, *enjoy*, *mind* etc.) and with the adjectives *busy*, *worth*, *like*
- the gerund with prepositions

(Kouluhallitus 1985, 73-75)[22]

In addition to knowing how to use verb forms, the 1985 curriculum called for knowledge of sentence formation and syntax: students erre to be able to form statements, commands and questions with the appropriate word order, to use question tags, to form compound and subordinate clauses, to create indirect questions and to use reported speech (Kouluhallitus 1985).

While the 1985 curriculum was exhaustingly detailed, the 1994 curriculum (Opetushallitus 1994) is written in a much more general way. The number of pages devoted to foreign languages is a good indicator of the change: while in 1985, English as the A-language alone took 22 pages (Kouluhallitus 1985, 61-82), in 1994, all modern foreign languages at all levels are discussed in only 6 pages (Opetushallitus 1994, 60-65). The 1994 curriculum does not offer any lists of grammatical forms to teach and, importantly, gives the same criteria for all A-languages. The goal is again to give the students tools for communication, but also to enforce their cultural identity and to expand their knowledge of the world. Furthermore, students were to learn skills and tools for participation in international cooperation and for work in international contexts. Another focus was for students to actively expand their skills and knowledge, increase their self-assessment skills and to take responsibility for their own learning (Opetushallitus 1994,

[21] The original includes examples of these structures, but they are omitted here.
[22] I have omitted the examples that were given in the original.

60). This curriculum also encouraged cooperation and integration with content teaching (Opetushallitus 1994, 62). Thus, the 1994 curriculum is much more open-ended and leaves room for teachers' individual decision-making. The curriculum adopts a more modern approach to teaching and learning, with more focus on choice and teachers' and students' own initiative (see also Pesola 2002). For a discussion on the effect of the curriculum the students in this study had been exposed to, see Section 7.2.1.

Although the focus in this study is on old curricula, a brief look at the curriculum applied later in the 21st century (Opetushallitus 2003; 2015), shows us that, again, no lists of grammatical structures are given. The focus is on communication, particularly intercultural communication. Multilingualism and multiculturalism are promoted[23], and more room is given to students' own initiative and interests. There is also a focus on the role of language in identity construction (Opetushallitus 2003, 100-101). This curriculum also links the goals with the CEFR (see Section 4.4.2) so that the intended target level in English as an A-language is CEFR B2.1, while it is CEFR B1.1 - B1.2 for all other A-languages as well as for English as a B-language (for A- and B-languages, see Section 5.2). In the most recent curriculum (Opetushallitus 2015), English is discussed separately from other A-languages, mainly because the intended target level differs from other A-languages. The curriculum now includes concepts such as metalinguistic skills and multilingual competence. The curriculum emphasises lifelong learning and students' own initiative: for example, students are now to self-assess whether their skills are sufficient for further study (Opetushallitus 2015). Furthermore, exploratory learning is encouraged, and students are to plan their own learning. For more details on the present curriculum, see Opetushallitus (2015).

5.4 Studies on Finns' skills in English

This section reports on research on Finns' skills in English, both when compared to other Europeans and when tested in nation-wide tests. This section also presents results from self-reported studies on Finns' skills and use of English. Section 5.4.1 focuses on holistic studies, while Section 5.4.2 reports on studies with a focus on specific aspects of English, often from a contrastive perspective.

[23] The CEFR (2001; see Section 4.4.2) also promotes multilingualism (Rothman, Cabrelli Amaro and de Bot 2013).

5.4.1 Holistic studies

This section focuses on studies that discuss Finns' overall skills in English, either in international or national tests or in self-assessment. These studies address both pupils, students and adults. I start the discussion with studies from the 1970s and then progress towards the present.

To begin in the past, Takala (2004, 258) reports on an international study from 1971. Here, the type of school pupils attended greatly affected their performance in tests in English. At that time, many Finns still studied German as the first foreign language and only started English at a later stage. Overall, 14-year-old students' level of English was modest but average in international comparison, while upper secondary school students had reached a good level in international comparisons (Takala 2004, 258).

Since the 1970s, similar international school-level comparisons have not been performed, but in the recent ratings of a private language school operating worldwide (EF 2015, 16), Finland was ranked in the fifth place. This result is based on scores from over 900,000 adults taking the school's online English test in 2014. The test-takers were self-selected and mainly included students and young adults. In the ranking, Finland is among the nine countries characterised with 'very high proficiency', which corresponds to CEFR level B2 (EF 2015, 61; for CEFR, see Section 4.4.2). However, the school maintains that while the ability to converse in English is widespread and many people have daily exposure to English, "many students do not develop an adequate level of academic English to pursue tertiary studies in the language" (EF 2015, 12; see also Sections 2.2 and 4.3).

National studies show that students' language skills have improved since the 1970s. Takala (2004, 265-270) believes this is partly because of greater access to English-language youth culture, increased exposure to English-language material, more emphasis on oral skills, particularly listening skills, and more versatile teaching methods. In tests given in 1991, teenagers in comprehensive school were able to do what students in the final grade of upper secondary school had been able to do in 1971. Takala (2004, 270-275) believes that the general level of Finnish pupils' English skills at the end of comprehensive school was fairly good in the early 2000s. The range, however, was wide, and although the best students were superb, the number of low-performers was, in his opinion, alarming.

Tuokko (2000) reports on a nationwide study in Finland, collecting information on ninth-graders' (15-year-olds') attainment of the goals in the comprehensive school core curriculum in 1999. Tuokko (2000, 31-32) reports that on average, pupils were successful in 64% of the tasks. About 14% of the pupils scored 'excellent' (more than 85% in the test), while 19% had weak skills (less than 45% in the test). In the grammar section, the most difficult topics were verb forms, articles and pronouns; the students succeeded with an average of 65%. The difficulty in verb forms mainly concerned the pupils' use of the present perfect in questions, the past perfect in statements, the conditional, the future and the passive in the past tense (Tuokko 2000, 59). In addition to school marks, the background factors which best explained success in the grammar section of the test were having an international penpal, active browsing of the Internet and reading English-language magazines (Tuokko 2000, 99). Tuokko (2007, 196-250) later created a method of indexing the skills of these ninth-graders on the CEFR scale (see Section 4.4) and argues that many of these students, about 40%, had reached level B1, and about 25% level B2.

However, some studies also report on differences within the population. In 1991, there was some geographical variation: students in southern Finland outperformed those in eastern and northern Finland. Pupils in bigger cities did a little better than those in the countryside, while the students' gender did not make a difference (Takala 2004, 267-270). According to Takala (2004, 272), regional and inter-school differences seem to have increased since, although scores in other school subjects have been fairly homogeneous. In 1999, Tuokko (2000, 35) reports that girls scored better than boys in all areas of the test, particularly in writing. Swedish-speaking pupils had better scores (74%) than Finnish-speaking students (63%).

Fairly similar results were provided in an extensive study by Sartoneva (1998) on Finnish adults' self-assessed language skills. In the study, 72% of adults reported that they can speak at least one foreign language, while 58% could communicate in at least two foreign languages. The most commonly spoken foreign language (66% of the adult population) was English (see also Leppänen et al. 2011). Again, there were differences in the population: women reported speaking more foreign languages than men, and young adults generally reported better language skills than older generations (Sartoneva 1998, 65-68). However, only 13% of the population (mainly highly educated young adults) had very good skills in English.

Many researchers (e.g. Leppänen and Nikula 2008; Leppänen et al. 2011; Meriläinen 2010a; 2010b; Takala 2004) argue that the role of English has become increasingly prominent in

Finland over the past few decades. The spread of the use of English in Finland has been so quick that Leppänen and Nikula (2008, 16; see also Meriläinen 2010a) argue that instead of English being a 'foreign' language, which it still was in the 1960s, '70s, and '80s, today English is almost comparable to a 'second' language in Finland, as a significant number of people use English on a daily basis (Leppänen et al. 2011, 16). An extensive survey on Finns' use of English in 2007 confirms that English plays a significant role in Finns' lives (Leppänen et al. 2011). Again, the best skills were among highly-educated young city-dwellers, while the older generations, people in the countryside and people with less education have lower skills. Most of the participants responded that they can use English at least moderately (Leppänen et al. 2011, 95-99). According to the study, most university-educated young people in Finnish cities have fully adopted English and use it in most aspects of life (Leppänen et al. 2011, 164-167). Thus, the participants in my study (see Section 6.2) should also be quite successful at using English.

Both self-reported and test-based assessments of Finns' skills seem to take very similar lines, with gradual improvement from the 1970s to today, and a similar tendency appears in European comparison. For example, recent Eurobarometer results show that 70% of Finns can have a conversation in English (European Commission 2012, 15). The score is higher than the European Union average, which is 38%; Finland is placed in seventh position out of 27 countries. Furthermore, 50% of Finns can listen to the news in English, 49% can read newspaper articles in English and 51% can communicate online (European Commission 2012, 31-37). Many Finns use foreign languages when travelling, watching television, on the Internet or at work, which is typical across Europe (European Commission 2012, 52), but 15% of Finns also report using foreign languages while studying something other than languages. This is a high percentage on the European scale.

5.4.2 Specific studies

A number of studies have focused on exploring Finns' skills in specific aspects of English, particularly with a contrastive focus. In many of these studies, the language skills of Finnish-speaking Finns have been compared to those of Swedish-speaking Finns. Some of these studies were conducted in the 20th century, but there are also more recent studies.

Many of the early studies conducted on Finnish learners of English focus on contrasting the English skills of Finnish-speaking and Swedish-speaking students, based on the fact that Swedish is typologically closer to English, while Finnish is structurally very different (see Section 5.1). Some of these contrastive studies applied error analysis, popular in the 1970s (see Section 3.2). Often, the result of these contrastive studies has been that because of the typological proximity between Swedish and English, Swedish-speaking Finns have an advantage in learning to use English. However, some of the studies conclude that the disadvantage that Finnish-speaking students have at the beginning is levelled out towards more advanced levels (e.g. Palmberg 1977). Thus, it seems to simply take Finnish-speakers a little longer to become proficient in English. While none of the studies presented below directly target verb use, they provide some useful information on Finns' skills in English.

Palmberg's (1977) results of an error analysis study with Finnish- and Swedish-speaking students at different levels of learning English show that Finnish-speaking Finns made many more errors in the use of articles, prepositions, verb-noun agreement and word order at the early stages of learning. The differences became less significant at higher levels but persisted in articles, prepositions and word order (Palmberg 1977, 88). However, there were also several individuals whose performance differed from the average patterns. Similarly, Sjöholm (1983, 178) found that Finnish learners of English make more errors with prepositions than Swedish learners. Ringbom (1977), who analysed spelling errors made by Swedish- and Finnish-speaking learners of English, shows that Finnish speakers made more spelling errors at the intermediate level, but they improved their skills to the advanced level, while Swedish speakers did not become better spellers as their skills became more advanced. Ringbom (1977, 106) explains that Finnish learners have to focus more on the difference between how words are spelled and uttered, while Swedish-speakers are already accustomed to some levels of mismatch in their first language.

An exception to the early studies is Lehtonen and Sajavaara (1983), who, instead of contrasting Finnish- and Swedish-speakers, studied whether Finnish learners of English need more time than native speakers of English for processing information in an error detection task focusing on articles in English. They found that although Finns needed more time to decide on the appropriateness of the sentences, the decision regarding correctness was similar (Lehtonen and Sajavaara 1983, 107). However, there were clear individual differences so that more proficient students had more automatised responses. In general, native speakers of English were more unanimous in their decision about the appropriateness of the sentences (Lehtonen and Sajavaara

1983, 109-111; see also Alanen 1997). Finnish informants displayed more disagreement and uncertainty, even with some sentences which all the native English-speaking informants found unacceptable. Another study that compares Finnish and native speakers of English is Crossley and McNamara (2011). In a study on advanced, university-level L2 writers from different language backgrounds, including Finnish learners, they note that Finnish students tend to apply tense and aspect repetition in similar ways and proportions to native speakers of English (Crossley and McNamara 2011, 280). Finnish writers had fewer words before the main verb than native speakers of English, and Finns' texts contained more negations than native writers' texts (Crossley and McNamara 2011, 278). However, none of these categories were generalisable to all learners.

In a study comparing Finnish students' reaction time measurements in grammaticality judgement to native speakers of English and German, Alanen (1997) found that more proficient students were faster and more accurate in their judgements, as explained in Section 3.3.2. In addition, she found that the fastest and the most accurate responses were given by students who "had spent either no time at all or less than one month in an English-speaking country" (Alanen 1997, 164). For similar results in my study, see Section 7.2.

There are also some studies contrasting Finnish and Swedish influence in the acquisition of vocabulary in English. Jarvis's (2000) study examined Swedish- and Finnish-speaking children mainly from lower secondary school to discover to what extent the children's L1 affected their L2 word choices and found that the L1 had more consistent effects on the outcome of lexical choice than variables such as age, length of instruction in English and task type (Jarvis 2000, 298). Another study by Jarvis and Odlin (2000) studied L1 transfer in spatial reference in Finnish-speaking and Swedish-speaking learners of English and found that Finnish-speakers had a greater distribution in the choice of preposition, while Swedish-speakers had greater intra-group consistency in their choice. Swedish-speakers do not show signs of Finnish influence, while some Finnish-speakers' English appears to be influenced by Swedish (Jarvis 2015).

In the 21st century, many studies exploring Finnish learners' skills in English have, again, mainly focused on the contrastive perspective. Meriläinen (2010a) provides an extensive study on Finnish upper secondary students' skills in English. Her study is based on the writing task in the matriculation examination, which is a high-stakes, nation-wide examination taken at the end of upper secondary school, typically at the age of 19 and usually after 10 years of studies in English (Ylioppilastutkintolautakunta). Using a large corpus with data on students' writing

from 1990 to 2005, she explored the appearance of deviant expressions, focusing on elements that might be L1-induced. Meriläinen's studies (2010a; 2010b) show that despite more communicative language teaching practices, Finnish students' language skills in 2005 seemed to be similar to those in 1990. She argues that "as measured through the frequency of transfer-induced grammar errors, the students' written English skills had not improved" from 1990 to 2005 (Meriläinen 2010a, 196). There is, however, positive development in students' skills in vocabulary, and it seems that L1 influence "is more persistent at the level of syntax than it is at the level of lexicon" (Meriläinen 2010a, 196).

In grammar, Meriläinen (2010a; 2010b) discovered five features where Finnish-speaking students crucially differed from Swedish-speaking students. The transfer-induced errors were found "in the student's deviant formation of the passive construction, the expletive pronoun construction, certain subordinate clause patterns, expressions for future time and prepositional constructions" (Meriläinen 2010a, 195). Each of these syntactic structures is formed very differently in Finnish and in English, and the study shows that these structures "are difficult not only for the weak but also for average and even good Finnish students to master" (Meriläinen 2010a, 114). However, the majority of the students formed these structures correctly. For example, for 69 deviant passives, there were 445 correct ones, and for 66 deviant structures with future time, there were 510 correct ones. Furthermore, when referring to future time, only structures with *'ll*, *will, shall* and (*be*) *going to* were explored (Meriläinen 2010a, 154), but not, for example, present tense structures that can have a future reference. Thus, a number of relevant structures were not investigated at all. However, verb forms were not the main source of problems for Finnish-speakers but rather word order, prepositions and articles.

As we have seen, many studies have focused on the differences between Finnish- and Swedish-speaking Finns in their acquisition of English, showing that Swedish-speakers have a distinct advantage. However, these studies have typically addressed aspects of grammar that are known to provide difficulty for Finnish-speakers. Finns' use of verb forms has rarely been addressed and is worth further inquiry. Furthermore, as the differences have tended to level out towards higher proficiency levels, it is important to study to what extent university students can use verbs in English.

5.5 Summary

The Finnish education system values language teaching and typically provides 10 years of studies in the first foreign language before tertiary level studies. English is the most commonly studied foreign language. However, given the extent to which English is used in Finland, it is not necessarily a 'foreign' language any more. Finns report on using English extensively and many young, educated Finns in big cities in southern Finland have good skills in English.

Studies comparing Finnish- and Swedish-speakers' skills in English indicate that Finnish-speakers tend to need more time to learn the way English is structured, while Swedish-speakers have an advantage because of greater structural similarities in grammar. Swedish is typologically much closer to English than Finnish, because Swedish and English are both Germanic languages, while Finnish is a Finno-Ugric language. At more advanced levels, the differences tend to be levelled out but may not disappear.

The upper secondary school curriculum has changed from a very detailed one in 1985 to more general ones in 1994 and in the 21st century. In 1985, the grammatical structures to be learned were carefully listed and divided between what students were to be able to understand and what they were to be able to produce. The more modern curricula focus much more on communicative competence and emphasise productive skills. Some studies suggest that while the move from a grammar-focused curriculum to a more communicative curriculum has improved Finns' skills in lexicon, the difficulties that Finns faced in grammar in the 20th century still persist.

Studies on Finnish students' skills in English show some evidence of transfer from Finnish, even at advanced levels. Finnish students seem to have difficulty with articles, prepositions, word order and some verb forms. These features seem to persist despite increased exposure to English and more extensive attention to communicative language use in teaching. Thus, greater awareness-raising in grammar would be beneficial.

6 The test, participants and methods

This chapter begins the empirical part and describes the test that was used to conduct the study, the participants and the methods that were applied. As explained in Section 1.3, this study aims to respond to the following research questions:

1. *What is the extent of variation in Finnish university students' use of English verbs in a fill-in-the-gap test, and what accounts for this variation?*
2. *What is the extent of variation in teachers' responses to the variation displayed by Finnish university students in their use of English verbs, and what accounts for this variation?*

To answer these questions, I asked Finnish university-level students to respond to a fill-in-the-gap test focusing on verb use and teachers of English to rate the responses. The focus of the study is to explore variation in these two participant groups and to outline Finnish students' general proficiency levels in the use of English verb forms. I answer the research questions in Chapters 7-10.

As explained in Section 1.2, there are gaps in research in these areas: variation in L2 English, Finnish students of English, university-level students, L2 use of English verb forms, Finnish teachers of English, comparing native and non-native raters and inter-rater variability. This study contributes to filling these gaps by studying Finnish university-level students' proficiency in the use of English verb forms, by analysing variation in the responses and by using different teachers as raters.

The study followed this outline: Students (see Section 6.2) were provided with a test at the beginning of a new course. This test was a fill-in-the-gap test focusing on verb forms (see Section 6.1). After I had received all the answers, I compiled a list of all the verb forms provided for each gap. I then recruited teachers (see Section 6.3) to assess the acceptability of these responses on a 4-point scale. The details and the rationale are provided below. The study followed the recommendations of the University of Helsinki Code of Ethics.

I begin this chapter with Section 6.1, which introduces the test that was used. Section 6.2 introduces the students and their background and Section 6.3 the teachers and their background. Section 6.4 outlines the methods that were used and the rationale for the choices that were made, and Section 6.5 discusses some issues that occurred during the coding of the data. Finally, the results of two pilot studies are briefly discussed in Section 6.6, because they provide relevant background for some of the choices made.

6.1 The test

Because the focus of this study is variation in language use, I needed a test that would provide a context and some limits to the variation that arises as it would otherwise become impossible to start rating countless expressions. This meant that free-writing activities were not suitable, nor were essays[24]. As I had become interested in this topic because of the variation in my students' responses to a particular exercise (see Section 1.1), I decided to use the same exercise that had first initiated my interest. As the exercise has several slots to fill in, focuses on verbs and is firmly contextualised, it appears a good choice to both attract but also limit variation so that it remains researchable. The exercise, here used as a test, is provided in Appendix 2. The test explores students' knowledge of discrete points (McNamara 2000, 14) in grammar, more specifically in verb forms.

The test is a 'traditional' fill-in-the-gap test, where students fill in the most appropriate form of the verb provided in brackets after each slot. The test is from an exercise book called *English Grammar in Use Supplementary Exercises* by Louise Hashemi and Raymond Murphy (1995). Written permission for using the exercise for this study was received from the publisher, Cambridge University Press, on 11 August 2003. The test has 107 slots to answer, and it tests students' skills in the use of English verb forms in the form of a detective story. First, there is a short explanation of the background of a murder. Next, there are four monologues ('witness statements') of the people involved, describing their actions at the time of the murder. This means that students were expected to provide written responses to a setting that comprises oral monologues. At the end, a closing dialogue is created between two police officers. Again, the dialogue is spoken, but provided in written form in the task. In the test, the first three slots had already been answered as examples, and the students proceeded from slot 4 to slot 110 to

[24] Many previous studies rely either on spoken data or on essays.

complete the story. Some sentences had no gaps to make it easier for the students to follow the story, but in most sentences, there was one or more verbs to fill in.

The full test set given to students consisted of four parts (see Appendices 1 and 2). The opening page of the test set explained the purpose of the study and gave the students the possibility to volunteer for an interview (page 1 in Appendix 1). The students began by filling in a questionnaire (page 2 in Appendix 1) with information about their language-learning background (see Section 6.2). Next, they were asked to fill in the test itself (Appendix 2). Finally, they proceeded to the self-assessment section (page 3 in Appendix 1), where the students responded to questions on their own feelings regarding how successful they felt they were in the test and their memories from English grammar lessons at school. The questions in the personal information questionnaire[25] and in the self-assessment section were in Finnish or Swedish[26], while the part with the test was, of course, in English. This study does not report on the responses provided in the self-assessment section nor the results of the student interviews due to limitations of scope and space, but they will be discussed in a later study.

After all the students had submitted their answers, I wrote a list of all the various verb forms provided as answers by these students. This list (Appendix 4) was then given to teachers of English (see Section 6.3) to determine how they would respond to the variation in the list and to indicate what they considered to be the best or acceptable answers and what they considered questionable or inappropriate choices in a specific slot. The recruitment letter that was sent to potential teacher participants is available in Appendix 3. In addition to rating the students' responses, the teachers also responded to questions about their background (see Appendix 5 and Section 6.3).

6.2 Student participants

The student participants in this study all studied at universities in the Helsinki metropolitan area[27] because access to these universities was easier for me. To be able to trace the factors affecting variability, I tried to find different student groups to participate. However, to keep

[25] The personal background questions were asked first, although some scholars (e.g. Dörnyei and Csizér 2012, 78) advise against this practice.
[26] The appendix is an English translation of the questionnaire.
[27] The University of Helsinki and the Finnish Academy of Fine Arts are located in Helsinki, while Helsinki University of Technology is based in the neighbouring city of Espoo.

variation under some constraints, a few prerequisites were included. Thus, this study focuses on students 1) who studied either at the University of Helsinki, Helsinki University of Technology[28] or the Finnish Academy of Fine Arts[29], 2) who had received the bulk of their primary and secondary education in Finland, 3) who had started studying English in primary school, 4) who had taken the matriculation examination in Finland and 5) whose first language was Finnish or Swedish. This allowed the focus to be on participants with a similar background. In addition, both A1 and A2 learners of English were accepted (see Section 5.2).

At the time of the test, students at the University of Helsinki were expected to complete a Reading Comprehension and an Oral Skills course or an exemption test in a foreign language in order to meet the degree requirements set by the university. The students were offered faculty-specific English for Specific Purposes courses so that they could focus on terminology and topics in their fields of study. Students who felt that their skills did not yet meet the level of the mandatory courses could take remedial courses first. In reality, however, some students took remedial courses after having completed the mandatory language requirements, and some students who would have benefitted from remedial courses prior to the mandatory courses did not take them. The Finnish Academy of Fine Arts had similar degree requirements for mandatory courses and offered English for Specific Purposes courses, but students in need of remedial instruction were asked to join the courses provided by the University of Helsinki. At Helsinki University of Technology, students were offered several mandatory courses to choose from. In addition, there were some remedial courses, which the students could take before attending the mandatory courses.

The test set was given to students in the academic year 2003-2004[30]. Their teachers served as my contact persons, since I did not teach any of these groups myself, and, as recommended by Mackey and Gass (2005, 34-35), I was not present in the situation where participants completed the questionnaires. Five teachers volunteered to help with this project, two at Helsinki University of Technology and three at the University of Helsinki, of whom one also taught at the Finnish Academy of Fine Arts. These five teachers were recruited on a voluntary basis; some other teachers were also asked to contribute but did not consent. This means that the sampling is, to some extent, convenience-based (e.g. Cohen, Manion and Morrison 2011; Dörnyei 2007; Dörnyei and Csizér 2012; Wagner 2015): the questionnaires were distributed by

[28] This university has since merged into Aalto University. However, I use the old name in this study.
[29] This university has since merged into the University of the Arts Helsinki. However, I use the old name in this study.
[30] Implications arising from the fact that the data are already old are discussed in Chapter 11.

the teachers who volunteered to do so, and thus the students that they happened to be teaching at that particular time constitute the sample, which makes the student population, at least in some respects, a purposive sample (see Vogt 2007, 81-82).

In contrast, the group of English Majors and the Remedial Grammar group were chosen specifically to be able to compare other groups to them, but they do not stand as control groups nor as criterion-groups (cf. Rasinger 2008, 42). The English Majors group and the Grammar group were recruited by approaching the teachers of these courses, who agreed to help. The students majoring in English Philology at the University of Helsinki took a mandatory course in English grammar, intended to be taken in the first year[31], while the Remedial Grammar course, which could be taken at any point and was optional, was intended for weaker students of any faculty. Thus, the students in this study represent various disciplines, ranging from students majoring in English to students from other academic fields taking a mandatory foreign language course and to students taking an optional, remedial course in grammar.

The courses the participants attended are listed in Table 4. The student participants at the University of Helsinki attended either an optional, Remedial English Grammar course or a mandatory English Reading Comprehension course at the Language Centre or a mandatory English Grammar course for students in the Department of English. Students at the Finnish Academy of Fine Arts also attended a mandatory English Reading Comprehension course, while students at Helsinki University of Technology attended either an optional, remedial Study Skills in English course or a mandatory Technical English or Advanced English for Everyday Use course. Unfortunately, their teachers failed to provide information on which student took which course, which means that these results are grouped into one category despite the fact that some students took a remedial and some a mandatory degree course. The test was distributed at the beginning of the autumn and spring term, i.e. in September-October 2003 and January-February 2004[32].

The students' own teachers gave the test set to each student in their classes without enquiring who would meet the criteria described above; this was done for classroom management reasons. The students were given the test and the questionnaire early in their course, either during the

[31] While most of the students probably did take the course in their first year, there might have been some students who had already studied longer at the university.

[32] As the students' year of birth is one variable, for the purposes of counting their age, it is assumed that everyone took the test in 2003. The potential error caused because of this system is a maximum of one year, which is not likely to create significant differences in the analysis.

first or the second meeting or as homework between the first and second meeting. For practical reasons, some groups were given the task as homework, while some groups worked on the test during class-time[33]. In both cases, taking part in the study was entirely voluntary. Intact classes[34] were approached, but any disinterested students had the possibility to opt out[35] without any sanction (for the importance of this policy, see e.g. De Costa 2015; Gass 2015; Mackey and Gass 2005). The students participated anonymously (for the importance of this policy, see e.g. De Costa 2015, 248) except if they volunteered for an interview (see Section 6.1), in which case their name and contact information was requested.

Table 4. The universities and the courses student participants attended

University	**Course name**	**Requirement**	**Time**
University of Helsinki	Remedial English Grammar	Optional	Spring 2004
	English Reading Comprehension	Mandatory	Autumn 2003
	Grammar (for English Majors)	Mandatory	Spring 2004
Finnish Academy of Fine Arts	English Reading Comprehension	Mandatory	Autumn 2003
Helsinki University of Technology	Study Skills in English	Optional	Autumn 2003 and spring 2004
	Technical English	Mandatory	
	Advanced English for Everyday Use	Mandatory	

I received a total of 353 questionnaires, of which 34 have been excluded from this study. Thirty-one students did not meet the criteria provided above, two students were disqualified as their answers had clearly been altered afterwards, and one student had received a faulty copy, with one page missing entirely. Of the remaining 319 students, 125 were from the University of Helsinki, with 35 students of English Philology (henceforth called English Majors[36]), 42 students from the Remedial Grammar group (henceforth called Grammar), 31 students from the Faculty of Arts (henceforth called Humanities[37]) and 17 students from the Department of Computer Science (henceforth called Computer Science). There were 12 students from the Finnish Academy of Fine Arts (henceforth called Visual Arts) and 182 from Helsinki University

[33] For potential limitations caused by this, see Chapter 11.
[34] This means that the full group was included in the study, and the group was a naturally formed student group. Whoever happened to take the particular course became a potential participant.
[35] I do not know how many students decided not to take the test at all, as the teachers did not keep a record of this. However, the teachers reported that this number remained very small.
[36] Of these 35 students, 30 actually have English as their major subject and five as their minor subject.
[37] This is to avoid confusion with the Visual Arts students; further, the Faculty of Arts is called the Faculty of Humanities at some universities in the world.

of Technology (henceforth called Technology). Table 5 lists the students groups and the number of informants.

Table 5. Student participants organised by student group

Student group	**Informant code numbers**	**Total number of informants (N)**	**% of all the informants**
English Majors	1-35	35	11.0
Grammar	36-77	42	13.2
Humanities	78-108	31	9.7
Visual Arts	109-120	12	3.8
Computer Science	121-137	17	5.3
Technology	138-319	182	57.1
total	**1-319**	**319**	**100**

The student population thus contained some students with presumably very good skills in English but also those with weaker skills. There were students who were entitled to study English Philology up to a Master's degree and who, therefore, could be assumed to have some expertise in English. The students taking a remedial course in English grammar, on the other hand, were all students of other majors and the course was entirely optional; the CEFR level (see Section 4.4.2) for the grammar course was B1. These students were, then, assumed to be weaker than average, and they had self-evaluated themselves as in need of more work in English before enrolling for their mandatory courses in English. In reality, however, the Remedial Grammar group also included stronger students and students who had already taken their mandatory courses but, nevertheless, wanted to revise grammar. Students taking the mandatory courses were presumed to be between the two extremes.

As explained above, students were also asked to fill in a background information questionnaire (see Appendix 1). The questions related to, for example, their majors, gender and marks from school, but also to their engagement with languages, English in particular. In addition to standard variables such as age and gender, questions were chosen to tap the students' previous success in English and their interest in languages in general. To better understand the background of the students participating in the study, Section 6.2.1 focuses on general background factors and Section 6.2.2. on questions specific to the students' engagement with languages.

6.2.1 Students' background information

In this section[38], I describe some of the students' general background factors: 1) gender, 2) the proportion of language vs. non-language majors, 3) the grade of starting to study English, 4) first language, 5) year of birth and 6) year of graduation from upper secondary school. Furthermore, whenever relevant, distributions are compared within the student groups, which provides information about the student population and explains which features they share. Section 6.2.2 discusses the students' exposure to English, their school marks and their level of engagement with English.

Of the 319 students who participated in the study, 171 were male and 148 female. Although the distribution of gender across the entire student population is fairly even (53.6% men and 46.4% women), this is not the case when the results are analysed across the student groups. There were more women than men in the English Major, Grammar and Humanities groups and more men than women in the Computer Science and Technology group. The only group where the distribution of gender was even was the Visual Arts group. This is visualised in Figure 10.

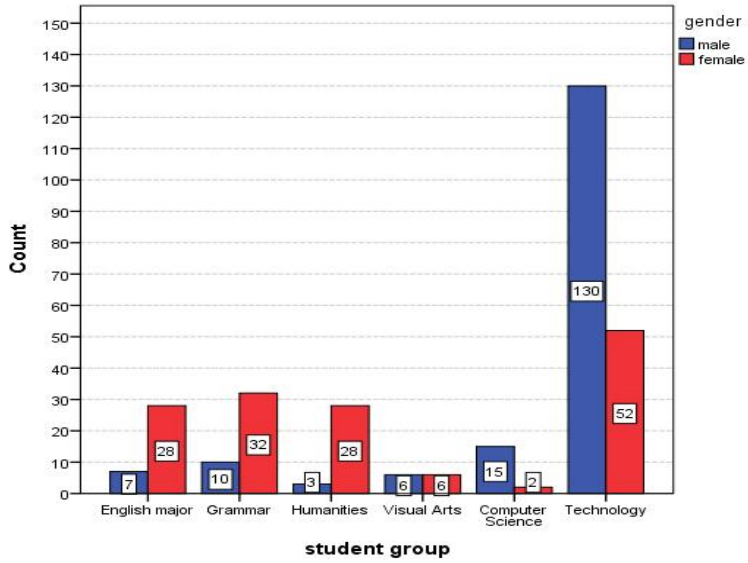


Figure 10. The number of men and women in the various student groups

[38] Throughout this section and the rest of this study, statistical test scores are given whenever they were run (cf. De Costa 2015, 251); when no inferential test scores are given, the results are based on analysing descriptive data. This is usually the case when the pattern is clear to the eye or when a test is not a feasible approach due to the small expected or observed values (e.g. below 5), or when inferential statistics are not appropriate.

Of the 319 students, 48 students (15%) majored or minored in a language-related subject, while 271 students (85%) studied a subject not related to languages. Obviously, all the 35 students from the English Major group had a language-related subject, but also nine students in the Grammar group and four students in the Humanities group majored in a language-related subject. These students included four students of Finnish and one student each of German, Spanish, Italian, Swedish/Hungarian, Russian, Nordic languages, Dutch, Slavic languages and General linguistics. The non-language related subject group comprised the Visual Arts, Computer Science and Technology groups together with the remainder of the Humanities and Grammar group students. Of the 48 students who studied a language-related subject, only 7 were men and 41 were women; in the non-language related group, 107 students were female and 164 male.

All the participants had started their English studies in primary school. The majority (79%) had started studying English in grade 3 (at the age of 9), while another large group of students (18%) had started English two years later, in grade 5. Only a few students had started English in grades 2, 4 and 6. The distribution is presented in Figure 11. Three students did not indicate when they had started their English studies. The students' gender distribution was even across the years of starting English, but while other student groups had fairly even distributions, Visual Arts students and English Majors were somewhat different. Of the 12 Visual Arts students, all but one had started their studies in grade 3. In the English Major group, somewhat surprisingly, there were proportionally more students who had started their English studies late: 31% of the English Majors had started their English in grade 5 instead of the more common grade 3 (69% of the English Majors). This means that almost a third of the students who studied English as their main or secondary subject had started with another foreign language, for example German or French.

In the student population, 298 students (93.4%) spoke Finnish as their first language, while 16 (5%) were Swedish-speaking (see Figure 12). Two students were bilingual[39] in Finnish and Swedish, and three were bilingual in Finnish and another language: Russian, German and Punjabi. Note that these are the students' self-reported first languages; all the students in the study spoke at least Finnish, Swedish and English (see Section 5.2). The distribution of the first language is not even across student groups: almost all of the Swedish-speaking students were

[39] Technically, most people in Finland are multilingual, with at least three languages (e.g. Cook 2016; Klein 1995; cf. Opetushallitus 2003; 2015). Here, however, I use the term bilingual to refer to people who self-identified themselves with two L1s. For various ways of defining bilingualism, see e.g. Brown and Larson-Hall (2012) and Ortega (2009); for research on bilinguals, see Foucart and Frenck-Mestre (2013).

in the Technology group (14 out of 16), and the Visual Arts and Computer Science groups consisted of only Finnish-speakers. Four of the five bilinguals were women, but in the Finnish- and Swedish-speaking groups, gender was evenly distributed.

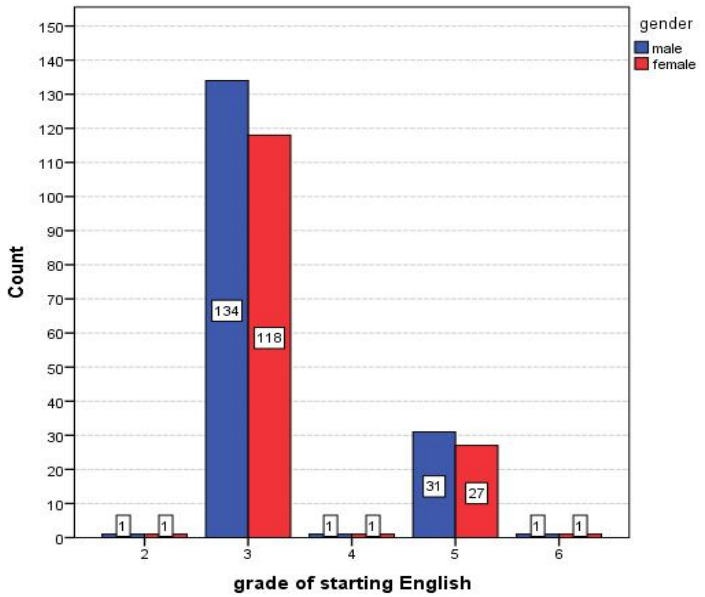


Figure 11. The grade when student participants started studying English at school

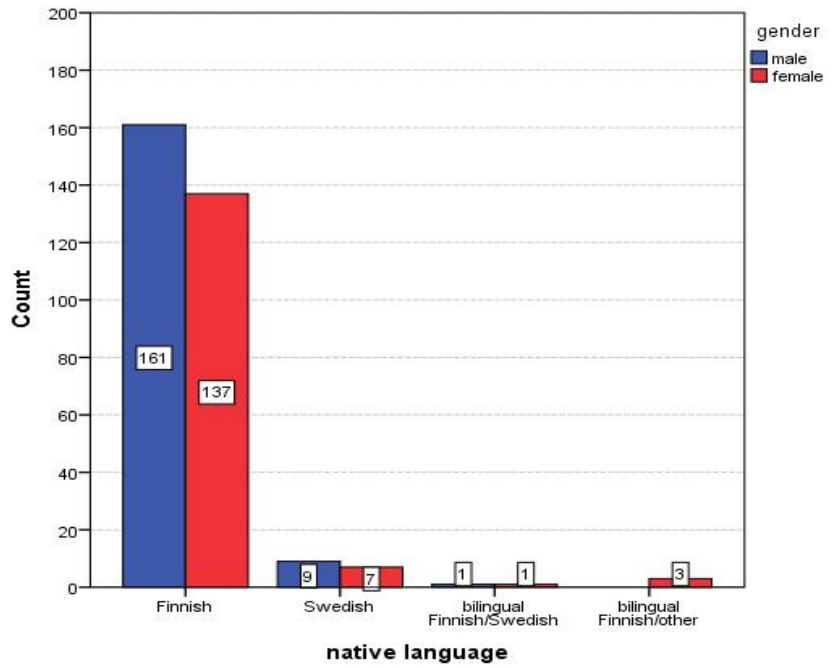


Figure 12. The distribution of the students' first language

The students' year of birth ranged from 1955 to 1985. Ten students had been born in the 1950s or 1960s, 112 in the 1970s and 197 in the 1980s. Thus, the students were 18-48 years old at the time of participation[40]. Figure 13 presents the range in the year of birth: most students had been born between 1976 and 1985, with the greatest peaks in the years 1981 and 1982. However, the distribution was not even in the student groups: there were proportionally more younger students in the English Major and Visual Arts groups. The English Major group, in particular, had more students who had been born after 1982 than any other group. This may be because the course they took was intended for first-year students, while there was no specification for when to take the course in the other groups. In this study, women were a little younger than men: the median year of birth for women was 1981 and for men, 1980, while the mode was 1982 for women and 1981 for men. Thus, on average, women were 22 years old, while men were 23 years old at the time of participation[41].

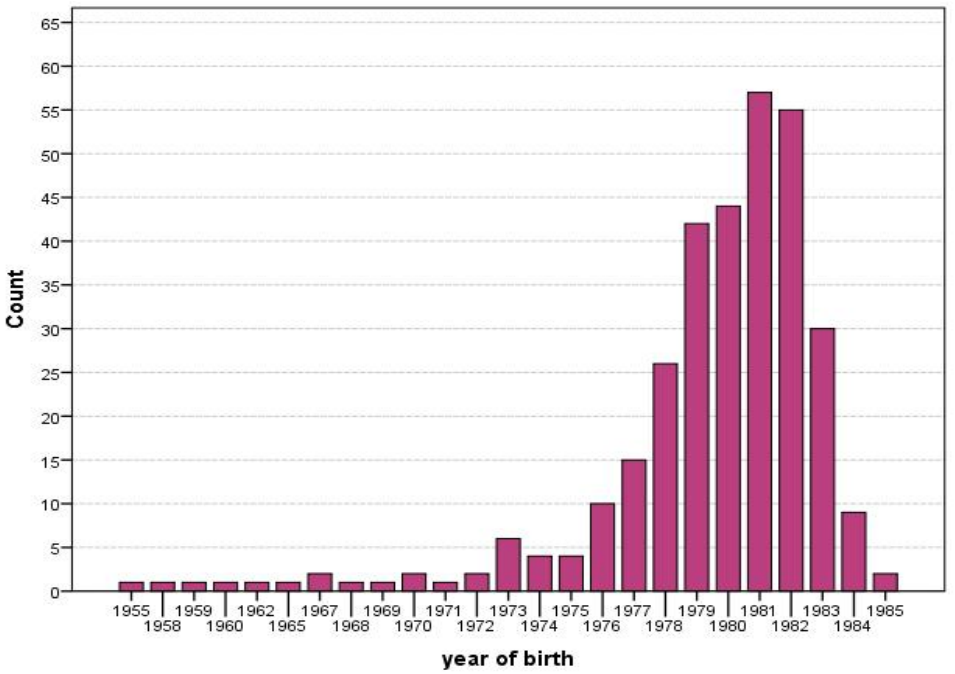


Figure 13. The distribution of the students' year of birth

The time from having taken the matriculation examination and having graduated from upper secondary school varied from 1975 to the previous spring (2003). However, the majority of

[40] For how the students' age was counted, see Footnote 32.

[41] One possible reason for this difference is the fact that in Finland, most men have to attend military service either before or during their university studies.

students had completed their matriculation examination one to five years before the time of participation (see Figure 14). This means that most students had been taught according to the 1994 curriculum, and only a minority had studied according to the 1985 curriculum (see Section 5.3).

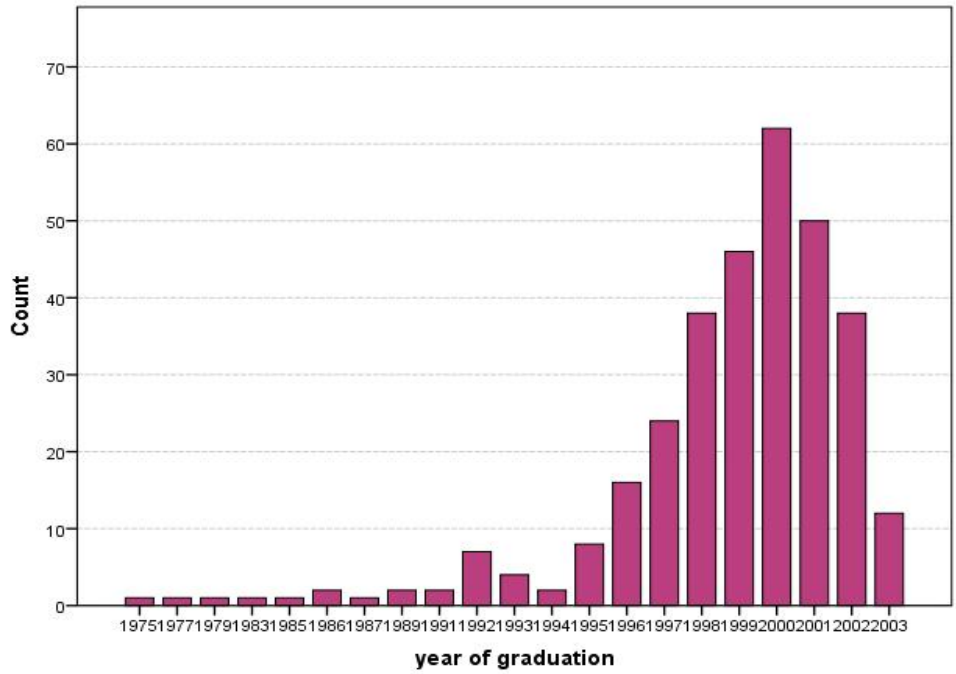


Figure 14. The distribution of the students' year of matriculation examination

To summarise, the participants were predominantly Finnish-speaking, had typically started studying English in grade 3 and were in their early twenties, having entered the university a few years after upper secondary school, but some participants were already much older and had had a long gap between upper secondary school and university. While women were overrepresented in language-major groups, men were overrepresented in the Computer Science and Technology groups.

6.2.2 Students' skills in English and engagement with languages

This section discusses the students' prior skills in English, focusing on their school success as well as their engagement with English before participation in the study. This section includes

details about 1) the students' marks in English at the end of school, 2) their marks in the matriculation examination, 3) whether they had studied more English after leaving school and before taking the course, 4) the length of visits to English-speaking countries, 5) potential other information the students thought might influence the test results and 6) the number of other languages they mastered. This enables me to compare the students' past proficiency and level of language engagement with their skills at the time of the test.

When leaving upper secondary school, one student had received the mark[42] 4 (fail) in English, while 133 students had earned the top two marks: 101 students (32%) had 9, and 32 students (10%) had the best mark, 10. However, students in the Grammar group were overrepresented amongst the students who had low marks, while the English Major group virtually consisted of students scoring 9 or 10, with only one student scoring 8. For the differences among student groups, see Figure 15.

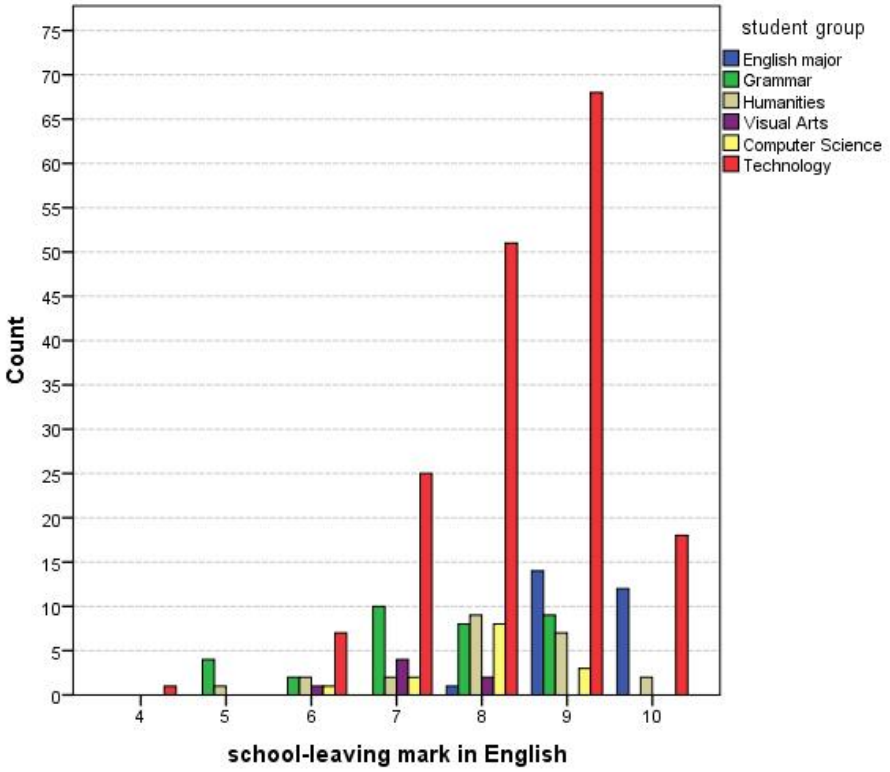


Figure 15. The distribution of the student participants' school-leaving mark in English

[42] School marks in Finland range from 10 (the best) down to 5, which is the lowest pass mark; 4 indicates fail. For more details, see e.g. Meriläinen 2010a.

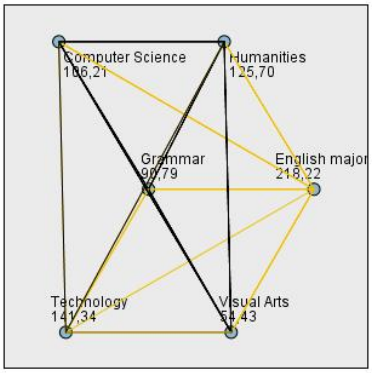
The Kruskal-Wallis test[43] shows a statistically significant difference in the distributions of the marks per student group ($H(5) = 54.62$, $p < .001$). The results of the pairwise comparisons are available in Table 6. As can be seen, English Majors stand out in particular, scoring a statistically significant comparison with each other group. The students' gender was evenly distributed across the marks (Mann-Whitney test[44], $U = 9340$, $p = .894$, $r = .008$), and Swedish-speakers' scores followed the Finnish pattern. For the distribution of first languages, the Kruskal-Wallis test yields $H(3) = 4.954$, $p = .175$, which indicates that the first language was evenly distributed along the marks, although the few bilingual students only scored at the top of the scale, with either 9 or 10.

A similar tendency is observed in the matriculation examination scores (for a discussion of the matriculation examination system and the marks used in Finland, see e.g. Meriläinen 2010a; for details about the examination, see Ylioppilastutkintolautakunta). Figure 16 shows that 123 students had been awarded the highest two marks, *laudatur* (46 students, 14%) and *eximia cum laude approbatur*[45] (77 students, 24%). Only one student had failed the matriculation test with an *improbatur*[46]. Again, the scores were not evenly distributed in the student groups: the Grammar students were overrepresented in scoring low, and the English Majors' scores were very high. The Kruskal-Wallis test gives a statistically significant score, with $H(5) = 77.178$, $p < .001$. In pairwise comparison, exactly the same pairs were signalled as statistically significant ($p < .05$) as with the school marks[47]. Again, women and men appeared evenly in the various categories ($U = 10323$, $p = .169$, $r = -.079$), while Swedish-speakers were overrepresented in the top categories when compared to Finnish-speakers, but the difference did not reach significance (Kruskal-Wallis, $H(3) = 6.679$, $p = .083$).

[43] The Kruskal-Wallis test is explained in Section 6.4.3.
[44] The Mann-Whitney test is explained in Section 6.4.3.
[45] This mark was introduced in 1996. Before 1996, *laudatur* covered this mark as well.
[46] In the Finnish matriculation system, you can compensate for a fail mark in one subject with good marks in some other subjects.
[47] The table is not reproduced here for this reason, although the precise levels of significance are not identical.

Table 6. SPSS scores[48] from the pairwise comparisons following a Kruskal-Wallis test, conducted with the Mann-Whitney test, examining the distribution of students' school-leaving marks per student group

Pairwise Comparisons of student group



Each node shows the sample average rank of student group.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
Visual Arts-Grammar	36,359	31,634	1,149	,250	1,000
Visual Arts-Computer Science	-51,786	35,191	-1,472	,141	1,000
Visual Arts-Humanities	71,267	32,816	2,172	,030	,448
Visual Arts-Technology	-86,913	29,319	-2,964	,003	,045
Visual Arts-English major	163,794	32,244	5,080	,000	,000
Grammar-Computer Science	-15,426	24,247	-,636	,525	1,000
Grammar-Humanities	-34,908	20,650	-1,690	,091	1,000
Grammar-Technology	-50,553	14,461	-3,496	,000	,007
Grammar-English major	127,434	19,728	6,460	,000	,000
Computer Science-Humanities	19,481	25,770	,756	,450	1,000
Computer Science-Technology	-35,127	21,138	-1,662	,097	1,000
Computer Science-English major	112,008	25,037	4,474	,000	,000
Humanities-Technology	-15,646	16,890	-,926	,354	1,000
Humanities-English major	92,527	21,571	4,289	,000	,000
Technology-English major	76,881	15,749	4,882	,000	,000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is ,05.

[48] Unfortunately, SPSS provides tables with commas instead of decimal points.

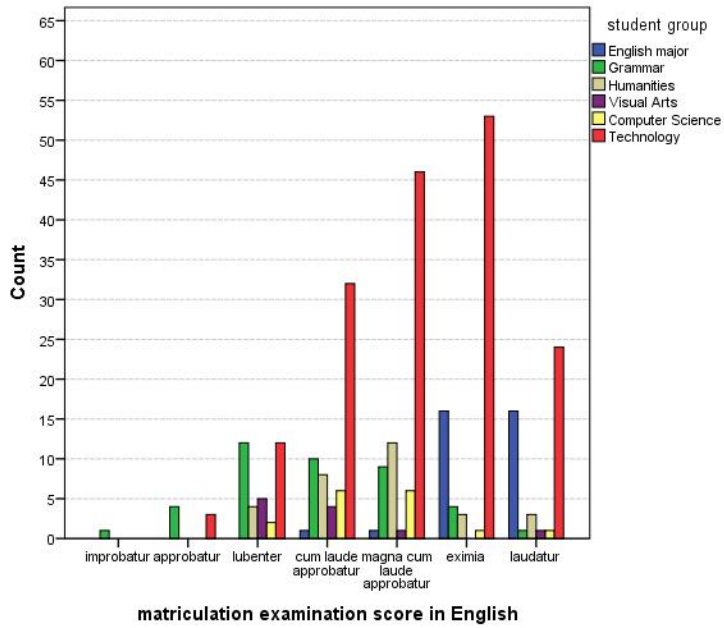


Figure 16. The students' matriculation examination score distribution per student group

Obviously, these results reflect the grouping as well: the students who scored well in the matriculation examination were more likely to be interested in applying to study English as a major, while students who scored low in the matriculation examination may have been more tempted to take a remedial course to improve their skills. Furthermore, Spearman's rank order correlation test[49] shows that the school-leaving mark and the matriculation examination mark correlate very strongly[50], $r_S = .804$, $p < .01$, which indicates that there is a strong association between the students' scores when marked by their own teachers and by the matriculation examination board examiners[51]. Interestingly, students who had started studying English later than the majority, i.e. in grade 5 or 6 compared to the more typical grade 3, scored somewhat more frequently in the top marks (Kruskal-Wallis for the school marks, $H(4) = 6.418$, $p = .170$ and for the matriculation examination, $H(4) = 12.766$, $p = .012$), which initially seems statistically significant for the matriculation examination marks, but pairwise comparisons show that only the pair '8 –*magna cum laude approbatur*' reached significance. There was no difference in whether the students had a language-related major or not (Kruskal-Wallis, $H(1) = 1.971$, $p = .160$).

[49] The Spearman's rank order correlation test is explained in Section 6.4.3.
[50] According to Salkind (2008, 82), correlations higher than .7 or .8 are rare in behavioural and social sciences, which means that the correlation here is very convincing.
[51] Note, however, that teachers preliminarily score their own students' test papers before sending them to the Matriculation Examination Board examiners.

Two-thirds of the students (212) had not taken any additional courses in English, while one third, 107 students, had taken one or more extra courses. These were, for example, optional studies either at school or at the university, student exchange periods or summer courses abroad. There was no statistically significant difference in the distribution in the student groups (Chi-square test for independence, $\chi^2(5, 319) = 1.602$, $p = .901$) nor between men and women ($\chi^2(1, 319) = 2.659$, $p = .103$).

Most of the students (262 students, 82%) had not visited an English-speaking country for longer than a month, while 47 students (15%) had spent from one to 12 months in a country where English is spoken as a native language. Only 8 students (2%) had spent more than a year in an English-speaking environment; see Figure 17. The student groups were, however, differently distributed: English Majors and Humanities students accounted for all the visits exceeding 13 months, and more women than men had spent time in English-speaking countries. Interestingly, even of the English Majors, as many as 22 (out of 35) students had spent less than a month in an English-speaking country. This may be because many of these students were first-year students and were perhaps planning to go on exchange later in their studies, as the degree requirements included some time spent in an English-speaking country.

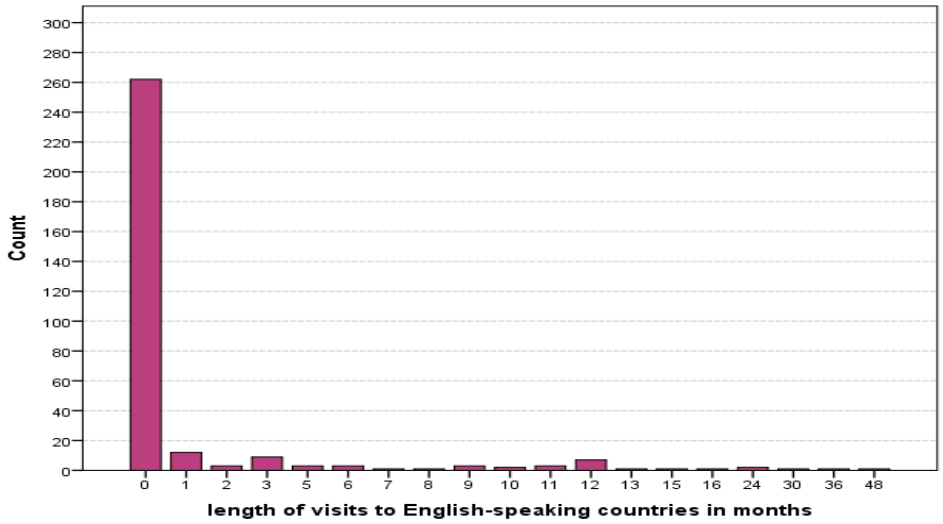


Figure 17: The length of student participants' visits to English-speaking countries in months

The questionnaire also enquired whether the students wanted to report on something that might have an effect on their results, although 84% of the participants (267 students) did not provide any such information. Of the 52 students (16%) who did, 12 students (4%) were engaged with what are traditionally called 'passive' (receptive) skills, such as reading literature in English,

while 20 students (6%) reported on 'active' (productive) skills, i.e. using or having used English actively and/or regularly in their lives. This included, for example, having an English-speaking partner, working in a company with English as the common language, or having lived in a country where the student had used English extensively. Twenty students reported on information that was not directly related to their English skills, such as having a headache, being tired while completing the test or having had a long break from studying English[52]. While all of these factors might affect the results, they have been categorised separately, as presented in Figure 18. It should be noted, however, that it is quite likely that some students who were actually engaged in either a passive or an active use of English did not come to think that what they did could be something that they should mention as being relevant in this particular context[53].

As can be seen in Figure 18, while men were more likely to report on a 'passive' use of English, no woman did so. Women and men did not differ from one another with respect to the other categories, but, interestingly, the distribution is uneven in student groups: only students in the Technology and Computer Science students reported using English passively, and if the Humanities and Visual Arts students reported on anything in this category, it was not related to language use. The English Majors either reported nothing or then mentioned active language use.

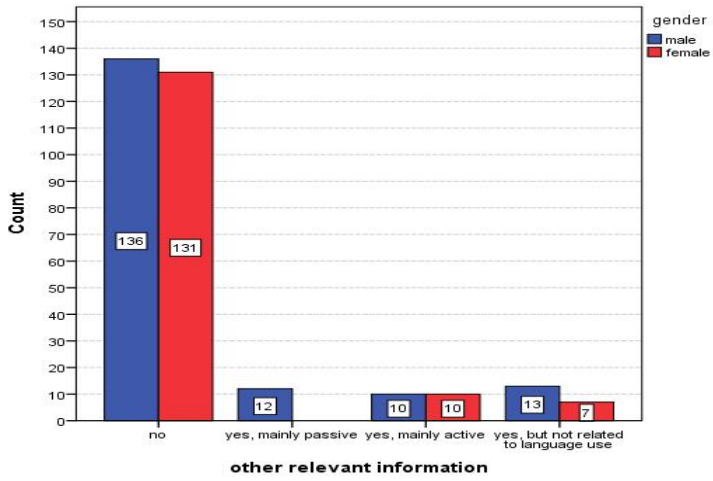


Figure 18: Student responses to the question on any other relevant information

[52] However, what the word 'long' meant was very individual. For example, one student mentioned half a year, another ten years.

[53] For example, activities such as the use of the Internet, programming or the consumption of media or films in English were only mentioned by five students, while it is likely that this kind of exposure was already frequent at the time of the test.

Finally, students were asked to list other languages that they can use in addition to English and their L1 (Figure 19). Seven students (2%) claimed to be unable to speak any other foreign language than English, while 86 students (27%) mentioned one other language, Swedish in most cases, and 143 students (45% of the student population) mentioned two other foreign languages[54]. The highest number of languages listed was six. The student group affected the distributions so that the two students who spoke either five or six languages were, perhaps surprisingly, both from the Grammar group, while the Visual Arts and Computer Science students spoke a maximum of two other languages in addition to English and their first language. The English Major group was somewhat overrepresented in the students who spoke three or four additional languages. Women were more likely to speak several languages than men (Mann-Whitney, $U = 16846$, $p < .001$, $r = .359$).

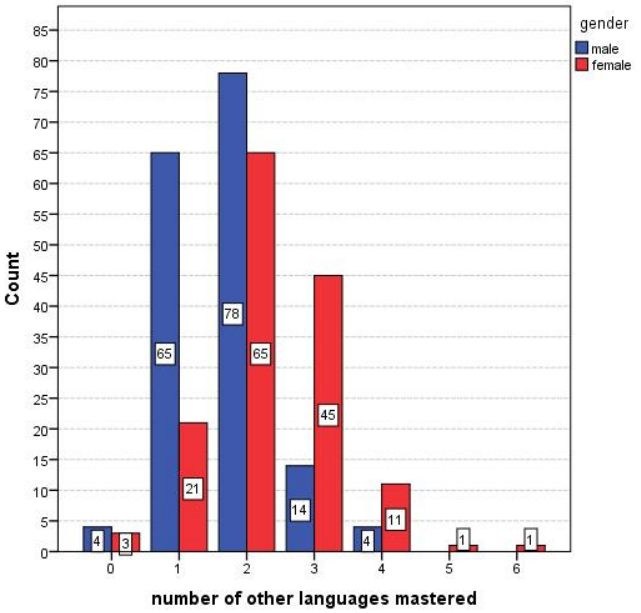


Figure 19: The number of other languages spoken by the student participants, split across gender

In summary, the student groups display partly different characteristics: the English Majors had had good marks at school and were more likely to have spent time in English-speaking countries, while the Grammar and Visual Arts groups had had lower marks at school and had not spent time in English-speaking countries. However, some individuals in the Grammar group

[54] While all of these students had studied Swedish at school, some Finnish-speakers did not mention Swedish here. It is possible that they believed that their skills were non-existent or that they did not think Swedish should be listed here as it is the second domestic language and studied by default.

could speak several languages and some of them, as well as some of the Humanities students, studied a language-related major. As explained in Section 6.2.1, the English Majors, Grammar and Humanities groups had proportionally more women than men in them, while the reverse was true for the Computer Science and Technology groups. How these background variables affected the test results is discussed in Sections 7.1-7.3.

6.3 Teacher participants

The participation of teachers was needed to assess the students' responses, and data from the assessments were used to explore how much variation there was in the teachers' ratings. for the purpose of convenience, the teachers were recruited among the staff at the University of Helsinki. The starting point was to recruit representatives of several varieties of English and from different backgrounds to explore the effect of different teacher identities on the responses and to explore whether L1 teachers of English rated the students' responses differently than L2 teachers of English, as has been indicated in some studies (see Sections 2.2 and 3.2.2). All English teachers at the Language Centre (a total of 20 in the academic year 2004-2005, both native English and native Finnish/Swedish speakers) and 20 teachers at the Department of English (all the native English speakers and a random selection of native Finnish/Swedish teachers[55]) were asked to contribute to the study in autumn 2004. However, as the completion of the task required a lot of time, the number of informants remained low (six teachers). Therefore, new recruitment rounds were undertaken at the Language Centre in 2008 and 2010, which yielded seven more teacher respondents[56].

The 13 teachers who serve as informants in this study all worked at the University of Helsinki at the time of participation, eleven for the Language Centre and two for the Department of English. Six of these teachers were native speakers of Finnish and seven native speakers of English, including two British, one Scot[57] and four Americans. There were six male and seven female teachers. Six informants had been born in the 1940s (the oldest in 1943), one in the 1950s, five in the 1960s and one in the 1970s (in 1973). At the time of their participation, the

[55] There were no other non-native English teachers than those with L1 Finnish or Swedish, and none of the Swedish-speaking teachers consented.
[56] Some informal interviews and discussions have also been conducted, but their contents will not be reported here due to their non-structured and informal nature.
[57] This is the self-identified nationality. In the subsequent analyses, this person is merged in the British category.

teachers were from 31 to 63 years old. Six informants had a Doctor's or Licentiate's[58] degree, while seven had completed Master's level studies. Four teachers had 1-9 years of teaching experience at the university level, five had 10-19 years of experience and four had over 20 years of experience. The Finnish respondents had lived in an English-speaking country from half a year to seven years, while the native English speakers had lived in Finland at least nine years, but some as much as 30 years; they had all arrived in Finland as adults, when they were from 23 to 54 years old. Table 7 lists the respondent details[59].

Table 7. Teacher respondents' background information

Code number	Native language	Years of university-level teaching experience	Years in English-speaking countries	Years in Finland
1	Finnish	5	1	*
2	Finnish	1	0.5	*
3	Finnish	22	1	*
4	Finnish	21	1.5	*
5	Finnish	26	1.5	*
6	Finnish	16	7	*
7	English (British)	10	*	10
8	English (British)	18	*	30
9	English (British)	22	*	24
10	English (American)	10	*	10
11	English (American)	8	*	9
12	English (American)	8	*	10
13	English (American)	15	*	16

* This piece of information is not given to protect anonymity.

These 13 teachers were given a questionnaire consisting of four parts: the recruitment letter with instructions (Appendix 3); the test, which was the same that the students had completed (Appendix 2); a list of all the various verb forms given by the students per slot in alphabetical order (Appendix 4) and a background information sheet asking about the teachers' background, language skills and attitude to grammar (Appendix 5). The teachers were asked to rate the extent to which they found the verb forms acceptable in the given context. They were asked to rate each form using the following scale:

[58] This Finnish degree roughly compares to an M.Phil. degree.

[59] Some variables, for example the year of birth, gender and level of education, are not provided here to protect the teachers' anonymity (cf. e.g. Phakiti 2015, 42; De Costa 2015; Cohen, Manion and Morrison 2011).

1 – this is the best alternative
2 – this is an acceptable alternative
3 – this is a questionable alternative
4 – this is an inappropriate alternative

The teachers were allowed to mark several forms as the best, acceptable, questionable and inappropriate. In rating the student responses, the teachers were not guided in any way in what definition of grammar or accuracy they should follow (see Chapter 3); actually, the instructions did not include the word 'grammar' at all. Rather, they were simply asked to evaluate the responses and to rate all the forms using the scale provided (the precise wording is available in Appendix 3), and the provision of a gradient scale allowed the teachers to decide how to apply the criteria.

However, the teachers were asked about their attitude to grammar in the background questionnaire, which they were to fill in after they had rated the verb forms. Interestingly, many teachers commented on their method of rating the students' answers in the background question "Any other information that might be important or useful", although this was not expected per se. Some teachers reported that they had developed a fixed system of reacting to different types of non-standard forms, such as spelling errors, unusual tense and verb change (e.g. using *phone* instead of *call*). The full responses to the question "How would you describe your attitudes towards learning, teaching and applying English grammar?" and to the (quasi-)question "Any other information that might be important or useful" are available in Section 9.3, where the teachers' self-reported approach to grammar is compared with their marking behaviour.

6.4 Methods of analysis

This section on methods is divided into three parts. I begin by providing the starting points for the study, including the reasons for the methodological and pragmatic choices that were made (Section 6.4.1). Next, I explain the choice of the test in Section 6.4.2, and finally, in Section 6.4.3, I discuss the statistical methods that were used.

6.4.1 Methodological and pragmatic choices

This section explains the methodological and pragmatic choices that were applied in this study. I originally intended to follow the mixed methods research approach, which involves "the combined use of qualitative and quantitative methods with the hope of offering the best of both worlds" (Dörnyei 2007, 20) and provides an improved understanding of abstract phenomena compared to either one approach alone (Rasinger 2008, 22; see also Johnson and Onwuegbuzie 2004, 14). However, the results reported on in this study are predominantly quantitative. This is because the wealth of data prevents me from discussing all the aspects I first intended, and some aspects will have to wait for a future study.

Although this study is mainly based on quantitative data, some elements of the original research idea with mixed methods remain. For example, while individual errors are interesting on their own, it is also interesting to know how typical and frequent such errors are (Chapter 7) and whether they follow any systematic pattern (see Section 8.5; cf. Chapter 4). While seeking common, perhaps generalisable tendencies, I also take an interest in individual cases; I am not only interested in how much variation there is but also how the variation displays itself. Furthermore, as can be seen from the teachers' data (see Chapter 9), it is not straightforward whether a verb form, even in a particular context, is 'right' or 'wrong': it would be counter-productive to use a binary division for a phenomenon that is complicated and forms a continuum rather than a dichotomy.

When analysing the number of non-standard forms that the students provided, depending on some of their background characteristics and the level of strictness applied (see Chapter 7), a rigorous quantitative approach is taken, including inferential statistical procedures. However, when exploring in detail the nature of the unconventional forms, I apply simple descriptive statistics, such as means or percentages. Additionally, some interesting cases that traditional quantitative methods might simply dismiss as outliers are discussed. My interest in outliers means that rather than ignoring them, they become one of the foci in the study. The decision to discuss outliers comprises some elements of case studies (for an overview, see e.g. Casanave 2015; Duff 2012; Richards 2011) in the spirit of what Dörnyei (2007, 271-272) would call either 'narrative profiling' or 'extreme case analysis'. The latter follows the principle of "examining the value of such unusual cases by identifying them by one method and further examining them

using the other" (Dörnyei 2007, 272). In this study, the choice of these cases is data-driven, not predetermined (cf. Johnson and Onwuegbuzie 2004; Rasinger 2008).

My study includes elements of survey research, which is research that uses questionnaires and interviews. Brown and Rodgers (2002, 16; see also J. D. Brown 2001; 2011) position survey research in the "common ground between the qualitative and quantitative approaches". The study also uses the perspective of triangulation, understood as "the attempt to understand some aspect of human behaviour by studying it from more than one standpoint, often making use of both quantitative and qualitative data in doing so" (Brown and Rodgers 2002, 243) or as the "attempt to map out, or explain more fully, the richness and complexity of human behaviour" (Cohen, Manion and Morrison 2011, 195). In this study, the 'three' in the triangulation can be seen, for example, in the sources of information (data triangulation: the students, teachers and the book used) and in the use of multiple sites (location triangulation: students from three different universities) (Brown and Rodgers 2002, 244).

There were 13 teachers involved in the study, and since they all expressed their opinions on acceptability, internal consistency measures or inter-rater reliability measures such as Cronbach's alpha (e.g. McKay 2006, 12; Phakiti 2015, 31-33; Révész 2012; Salkind 2006; 2008) could have been calculated. However, since the purpose of the study is neither to come to an agreement as to what the 'right' answer is nor to produce a perfect test, inter-rater reliability in the form of indices is neither needed, calculated nor reported; rather, I take an interest in the variation that exists in the responses. However, since the acceptability of student responses is analysed, I explore the students' results in the light of inter-rater agreement at several levels (see Chapter 7), which examine the phenomenon extensively and in detail and address both inter-rater reliability and the internal consistency of the test. As explained in Section 3.3, this study has little to do with grammaticality or acceptability judgement studies despite the use of the term 'acceptable' in the rating scale. No analyses often undertaken in language testing research, such as item difficulty analyses (for an overview, see Read 2015), were performed since I do not attempt to standardise the test in any way.

This is a cross-sectional (correlational) study: the questionnaire and the test provide a snapshot of the students' skills at a given point in time (e.g. Dörnyei 2007; Field 2013; Phakiti and Paltridge 2015, 12; Rasinger 2008; Whong and Wright 2013). The study tests students' explicit knowledge (Ellis 2006, 95; see also Andrews 2007, 13-14; Storch 2015, 349-350) of the use of verb forms, since the students were aware of the fact that they were supposed to produce the

suitable form of the verb provided and since all of the items tested were verbs. Furthermore, the test examines the students' knowledge of English verbs, not their general skills or ability in English (for the difference, see e.g. Borg 2015; Read 2015). The students' skills are discussed both on the basis of their general proficiency in the use of verb forms and their responses to single slots.

6.4.2 The test

The study draws on data from a test employing the fill-in-the-gap format (also called a cloze[60]). As Storch (2015, 350) notes, such a format is a traditional task for researching explicit knowledge in grammar, and Tremblay (2011, 364) finds cloze tests a feasible way of researching students' skills. The fill-in-the-gap test in this study was a paper-and-pencil test, but the responses were not fixed, i.e. they were not selected from a limited list (for types of tests, see McNamara 2000). As a test type, a fill-in-the-gap test is a discrete-item test, where "individual components of the learner's knowledge" (Thornbury 1999, 141) are measured. As such, it is considered effective, practical, reliable and valid for testing a limited area of learners' knowledge (Thornbury 1999, 141-143).

There is some research on the validity of fill-in-the-gap tests. For example, Tremblay (2011, 344) employed a cloze test to examine whether it matched the results of other ways of testing students' skills. She cites previous studies supporting the fact that cloze tests are internally consistent, reliable and discriminable, although ultimately this, of course, depends "on the extent to which these tests are tailored for the targeted population" (Tremblay 2011, 345). She further argues that cloze tests are useful as they are practical, easily created and modified and quick to mark. In her study, there was a significant relationship between cloze scores and broad proficiency estimates derived from the participants' language background (Tremblay 2011, 357). Her study included a "bank of acceptable answers", created by the author herself after consulting dictionaries for unclear cases and by allowing for spelling and agreement errors as long as pronunciation was not affected (Tremblay 2011, 351). While my study also includes the concept of an 'acceptable' answer, the method I used is different, for my study includes 13 teachers as raters (see the criticism towards single raters in Section 3.3.2).

[60] I use the term fill-in-the-gap, because Thornbury (1999, 145) notes that true clozes have every *n*th word deleted, which means that the item needed might be any part of speech, while in this test all the gaps required verbs.

I used the exercise exactly as it is provided in the book (see Section 6.1), with no attempts to withdraw slots that are ambiguous, although this procedure is often suggested in designing a test (e.g. Mackey and Gass 2005, 96). This is because the focus of this study is to research what kind of variation occurs and is prompted by the test as it is, not to produce a perfect test. Thus, I did not have specific target structures in mind that should have been controlled for (Mackey and Gass 2005, 104-5); nor did I attempt to create any.

6.4.3 Statistical methods

This study relies on non-parametric tests for several reasons. First of all, large parts of the data are not normally distributed (e.g. the Shapiro-Wilk test of normality[61] for the number of errors as a dependent variable gives $W = .763$, $p < .001$ and the Kolmogorov-Smirnov test $D(319) = .210$, $p < .001$), which means that parametric tests should not be used. Many of the distributions in the data are both skewed and have kurtosis, which means that they deviate from normal; this is partly because there are several outliers. Moreover, the variables are typically not linearly related and not additive, and Q-Q plots show clear deviations from normality. Even taking the central limit theorem (e.g. Cohen, Manion and Morrison 2011; Field 2013) into consideration and knowing that large sample sizes begin to approach normality, it is likely that only the Technology group might be big enough to give normally distributed scores (however, the Shapiro-Wilk and Kolmogorov-Smirnov tests indicate that the distribution is not normal even for this group). Furthermore, since I am interested in the outliers, I did not want to manipulate the data by trimming, winsorising or transforming it (Field 2013, 196-210; Pallant 2013). Given that the existing non-parametric tests are sufficient for the purposes of this study, there is no need to bootstrap the results, either. Additionally, when discussing the variation in teachers' evaluation, the study does not employ more refined statistical analyses such as Facets, used by e.g. Huhta et al. (2014) and Kondo-Brown (2002), because the aim of the study is not to arrive at inter-rater reliability but to discuss variation as it exists.

For the quantitative analyses, SPSS software (IBM SPSS Statistics 22 and 24) was used. For help with the computation and interpretation of quantitative data, I consulted Dörnyei (2007), Field (2013), Laerd Statistics (2013), Pallant (2013), Rasinger (2008) and Salkind (2008). The results were analysed with non-parametric tests, including the Mann-Whitney test, Kruskal-

[61] In addition to the two tests of normality, I have followed Pallant's (2013, 59-60) and Field's (2013) advice and also verified skewness and kurtosis from histograms.

Wallis test and Spearman's rank order correlation test (see Table 8). The Mann-Whitney test examines the "differences in the ranked positions of scores in different groups" (Field 2013, 224) when there are two groups. Thus, it is the non-parametric equivalent of the independent-samples *t*-test. The Kruskal-Wallis test is based on ranked data and examines whether the population medians of a dependent variable are the same across more than two groups. It tests whether "multiple independent groups come from different populations" (Field 2013, 236) and is similar to its parametric counterpart, the one-way between-subjects ANOVA. Whenever a Kruskal-Wallis test was used for the initial comparison between groups, it was followed by the Mann-Whitney analysis for the multiple comparisons. For any post hoc analyses, the Bonferroni correction was always applied, as recommended by e.g. Kline (2004, 71), Pallant (2013, 243) and Salkind (2008). Spearman's rank order correlation, also known as Spearman's rho, is often used to analyse two sets of ordinal data (Brown and Rodgers 2002, 170) and is the non-parametric counterpart to Pearson's correlation (e.g. Pallant 2013, 134).

Table 8. The main statistical tests used in this study (adapted from Pallant 2013)

The non-parametric test used	**Purpose of the test**	**The parametric counterpart**
Mann-Whitney	Is there a difference in the ranked medians between two independent groups on a continuous measure?	Independent-samples *t*-test
Kruskal-Wallis	Is there a difference in the ranked medians of a dependent variable across three or more groups?	One-way between-subjects ANOVA
Spearman's rank order correlation	Is there an association between two ranked variables, and what is the strength and direction of the relationship?	Pearson's correlation

I am aware of the fact that non-parametric tests "tend not to be as powerful" as parametric tests and that they may "be less sensitive in detecting a relationship or a difference among groups" (Pallant 2013, 116; see also Kline 2004; Salkind 2008). Nonetheless, Pallant (2013, 213) suggests using non-parametric tests when the test population is not a random sample of the entire population, and Vogt (2007, 67-68) recommends non-parametric tests when the data are not normally distributed, when they are irregular or skewed or when sample sizes are not equal. He also argues that "in some circumstances, nonparametric tests can actually be more powerful, especially when the values of the variables are not normally distributed" (Vogt 2007, 69). Since data transformation is considered somewhat controversial in any case, the data were analysed with non-parametric tests in this study.

For some simple variables, such as gender and language vs. non-language majors, the codes existed before the analysis, while for some more complex variables, coding was decided on during the process based on what emerged from the data. This type of coding for the background variables was conducted *in vivo* (coding that emerged from the analysis when patterns were observed and established; see e.g. Baralt 2012, 230-231) because the nature of the responses could not be predetermined.

6.5 Issues with the data

Unfortunately, a few issues took place during data collection and analysis that may influence the results to some extent. These issues concern the discrepancy between the lists of verb forms and the fact that some of the participating teachers did not fully follow the instructions they were given. However, while this is regrettable, luckily these issues do not affect the results in a serious way and remain marginal.

The list given to the teachers to rate differs slightly from the list of actual student production. Luckily, however, the effect the difference has on the results is minimal, and there are various reasons for why this happened. First, the list given to the teachers was drawn up before students who did not actually qualify were excluded; for this reason, there are a few forms that such students used in the teachers' version, although no student in the study itself used such forms. Second, the students responded to the test and the questionnaire by hand. It was difficult to read some people's handwriting, but I have done my best to do the students justice. When potential spelling errors were spotted, the relevant letter or letters were always compared with other instances of the same letter to verify that they were not misinterpreted. In cases that remained unclear, the students were always presumed to have spelled the word correctly. However, if misspelling seemed evident, the word was recorded misspelled even in cases where the intended response would have been easy to guess, as in *lalked* for *talked*. This was done to be systematic in the coding. The student response sheets were examined twice, first for compiling the list for teachers and later for the actual analysis. Occasionally, if the handwriting was ambiguous, my decision on what letter(s) the students had used seems to have changed the second time. To verify transcriber and intra-coder reliability (e.g. Révész 2012, 216; Gass 2015), all lists, files and computations have been double-checked, some even triple-checked, after the original data entry, coding and analysis.

The third reason for some discrepancy between the student and teacher lists is that the computer program used (Microsoft Word) first kept autocorrecting some non-standard forms (e.g. *did'nt* was changed to *didn't* automatically), despite attempts at switching off the automatic correction function. This means that some unconventional spelling provided by the students is missing from the teacher list but exists in the student list. Fourth, during the first round of listing, double answers (where a student gave two answers to a slot, e.g. *am talking* / *was talking*) were listed as two separate items (e.g. as *am talking* and *was talking*), while during the second round, double answers were listed as one unit, with the two forms separated with a slash (/). The same system was followed if the student's response contained brackets (e.g. *takes normally* (*a*)). Fifth, the list from the second round also includes the cases where no answer was given. Sixth, some answers had accidentally not been included during the first round.

However, the above shortcomings have a minimal effect on the results. The vast majority of responses that were not included in the teachers' list were given by one student only. This is 149 forms, of which 37 were forms where there was a double answer with either a slash, as in *visited* / *had visited,* or brackets, as in (*had*) *had*. There were also instances of no answer in 57 slots, with 1 to 44 students not supplying an answer, resulting in a total of 207 cases of no answer, and 9 forms with 2 students supplying the same answer that were not rated by the teachers. Similarly, of the verb forms listed on the teachers' version but not included in the actual student responses, 150 out of 211 have been rated as inappropriate by all the 13 teachers, and in 32 cases, 12 teachers have rated the form as inappropriate. This leaves 29 cases where 11 teachers or fewer found the form inappropriate, but some teachers also found the form the best, acceptable or questionable. Since the number of these problematic cases is small and mainly affects individual students, and not systematically the same student, these accidental omissions do not bias the results. For a discussion on sources of error in coding, see e.g. Révész (2012, 204), who grants that it is "almost impossible to eliminate errors completely".

Although the students were asked not to use any material to help them when completing the test (see Section 6.2), some were given the test as homework, and there is no guarantee that they did not, for example, consult a grammar book or discuss their choices with someone. However, as the students would gain nothing from disregarding the instructions for the test and as no student had a perfect score, it is very unlikely that they did so. In addition, one of the teachers chose to go through the test together with the students, asking them not to change their original answers after learning the suggested correct alternatives. Nonetheless, she was not able to control that this did not take place. Two students were disqualified from the study as in their

cases it was clear the responses had been altered (see Section 6.2). It is possible that some other students did so as well without this being as evident as in the abovementioned cases, but again, students would gain nothing from doing this and therefore it cannot be considered likely that they altered their responses. Intentional silliness (cf. Dörnyei 2007, 204) is unlikely as the students were no longer teenagers.

Unfortunately, the fact that teachers at Helsinki University of Technology did not keep the test sets given to students taking different courses separate may bias the results, because there were both students from mandatory and optional, remedial courses. However, the Technology group scored well, but it would have been interesting to know whether the students from remedial courses at the University of Helsinki and Helsinki University of Technology would have provided similar scores and whether the low-scoring students in the Technology group were from a remedial course.

6.6 Pilot studies

Before I began this study, I conducted a study on a related topic as a part of my teacher training programme in 2002 with 58 students and four teachers at the University of Helsinki (Pesola 2002; for the importance of piloting, see e.g. Wagner 2015, 89-90). The results of the study encouraged me to investigate this topic in more detail. It also serves as a pilot study, although it was not originally created for such a purpose. Another pilot questionnaire was distributed among students at the University of Helsinki and Helsinki University of Technology in the spring of 2003, with the sole purpose of testing different language versions and the wording of the questionnaire and its suitability for this study. This test set was distributed to 38 students.

It was discovered that when the students were asked to complete the background questions in their first language, they gave more detailed information, perhaps because they did not need to focus on deciding which words in English would correspond to their intended meaning. Some researchers argue that the comments become more extensive when respondents are allowed to answer in the language they are most comfortable with (e.g. McKay 2006, 53), and Dörnyei and Csizér (2012, 79) believe that the quality of responses improves when the respondents answer in their first language. It was, therefore, decided that in this study, the questionnaire section should be written in the respondents' first language. In addition, on the basis of the pilot tests, a number of questions in the background section were edited to form the present

questionnaire, as a few questions were misinterpreted (for a discussion on the difficulties posed by informants understanding questions in deviant ways, see e.g. Piispa 2006, 152-154).

The first pilot study (Pesola 2002) showed that Finnish students who had a command of another foreign language in addition to English made fewer errors in English than students who did not have good skills in other languages. Furthermore, students who were motivated to study English and who felt that they had received good instruction at school scored higher than other students. In addition, students seemed to be good at making realistic assessments of their own skills, and their school marks were good predictors of success in the test (Pesola 2002).

In the first pilot study (Pesola 2002), the main sources of error were 1) the progressive vs. the simple form, 2) wrong tense, 3) gerunds and infinitives, 4) active and passive forms, 5) missing and added words and 6) word order. It was also observed that some students were not systematic in their responses. For example, they might at times produce the past progressive and at other times the past simple in slots that would require the past progressive. Some students barely used any progressive forms at all. The second pilot study (2003) was not published, because it was used solely for verification purposes and on testing whether the questionnaire should be in English or in Finnish/Swedish. As explained above, it was found that it was easier for students to respond when the background questions were in their first language. I also followed Dörnyei and Csizér's (2012, 79; see also Mackey and Gass 2005) advice on collaboration to verify that the two language versions were compatible with one another and consulted two bilingual teachers of Swedish in relation to the wording of the Swedish version.

6.7 Summary

This study investigates how variation is manifested in Finnish students' use of verb forms. The study was conducted in the form of a fill-in-the-gap test and a background information questionnaire. The participants filled in 107 slots using the verb provided in brackets. The test was conducted at the beginning of a course the participants took as a part of their studies. The students filled in the test in English and responded to background questions in Finnish or Swedish. The test examined students' explicit knowledge of verb forms in the context of a detective story.

The participants studied at three universities in the Helsinki metropolitan area in 2003-2004. They comprised English Majors, students in the Faculty of Arts, students majoring in Computer Science and students taking a Remedial Grammar course at the University of Helsinki. There were also students from the Finnish Academy of Fine Arts and Helsinki University of Technology, and in total, there were 319 students. The student sample is partly convenience-based, partly purposive and partly specifically chosen. All the students had received their primary and secondary education in Finland, passed the matriculation examination and spoke either Finnish or Swedish as their first language. They had all started studying English in primary school.

The students' responses were listed in alphabetical order and teachers were recruited to rate the responses on a scale of four: the best answer(s), acceptable answers, questionable answers and inappropriate answers. Thirteen teachers of English from the University of Helsinki participated in the study. This included six Finnish, three British and four American teachers. They also completed a background information questionnaire. All the teachers had experience teaching English at the university level in Finland.

The data were analysed with SPSS using quantitative methods. The statistical tools used were non-parametric and included the Mann-Whitney test, the Kruskal-Wallis test and Spearman's rank order correlation test. The data were used to explore both students' proficiency levels and variation in their responses. Two pilot studies were conducted before this study. Their results indicated that the main sources of verb error for Finnish students were with aspect, tense, gerunds and infinitives, active and passive forms and word order. Some errors were caused by dropping or adding words.

7 Students' overall proficiency

My study focuses on the variation that is evident in students' production, not on pinpointing errors. However, in order to discuss overall proficiency, I have to use the concepts of "correct answer" and "error", since it is not possible to examine the extent of variation without knowing which forms are considered acceptable and which are not. I need a term for discussing forms that deviate from the standard or from the acceptable forms. In this study, I use the term "error" in conjunction with exploring the students' overall skills. Such labelling is mainly applied for statistical purposes so that overall trends can be observed and discussed. The criteria for acceptance come from the teachers who participated in rating the responses (see Section 6.3) and from the book key (see Sections 6.1 and 7.1.2). In Chapter 8, when I discuss the forms students provided in more detail, I am more interested in the extent of variation and also use terms such as "unconventional", "non-standard" or "unusual" to refer to verb forms that do not comply with the expected form.

In this study, errors are primarily understood as the forms rated inappropriate or questionable by the teachers. Their number per student is calculated using four different criteria: the most lenient, moderate, conservative and the strict. The students' background information is compared against these scores. The precise nature of the errors, then, is discussed in Chapter 8 from different perspectives, including a focus on particular individuals and with a particular interest in both common trends and unusual cases, which traditional quantitative studies might dismiss as outliers. The variation in teachers' assessment is discussed in Chapter 9.

The discussion of the results regarding students' overall proficiency begins from the range of variation by focusing on general trends and the extreme ends of variation (Section 7.1). This section also introduces the book key and its suggested correct answers as well as the overall trends in teacher variation in rating the students' results. The results are explored through analysing the overall scores across the student groups and the students' background factors. The overall proficiency level is then given at the most lenient level (Section 7.2), after which it is recalculated using a moderate, a conservative and a strict approach (Section 7.3). These three sections aim at providing the scope of variation on the full sample level. Throughout this

chapter, it is useful to refer to Appendix 2, where the test is available. However, wherever practical, the immediate slot context[62] is provided in the text.

7.1 General trends

The aim of this section is to provide overall statistical data on the extent of variation and general trends in variation in this study. I begin the discussion of variation in the students' responses with the extreme ends of variation in Section 7.1.1. Next, I discuss the results in relation to what the writers of the textbook expected to be provided in Section 7.1.2. Finally, the nature and extent of teacher consensus is outlined in Section 7.1.3.

7.1.1 The extremes of variation

The total number of verb forms supplied was 34,133, if we include instances of no answer, which numbered in 205. Students gave a total of 1,526 different verb forms to the 107 slots in the test. The full list of the verb forms is provided in Appendix 6. I begin the discussion with the range of variation in the responses, measured first in how many different forms there were per slot and then in how many students provided exactly the same form per slot. Figure 20 illustrates these extreme ends of variation.

The smallest number of variant forms given per slot was 4, with slot 55 (**So I ___ (walk) home again**). This particular slot is, then, the one with the least variation, if seen from the perspective of the number of alternative forms suggested. Here, 312 students out of 319 provided the form *walked*. There were three other forms given: *walk* by four students, *went* by two students and *had to walk* by one student. The teachers unanimously agreed that *walked* was the best alternative and that *walk* was inappropriate, but they had different opinions on the acceptability of *went* and *had to walk*.

[62] I have, however, systematically dropped the slot numbers from the slot quotations. In the original, they were provided immediately before the slot, as e.g. in "I (59) ___ (call) at the Sterns' house at nine-fifteen."

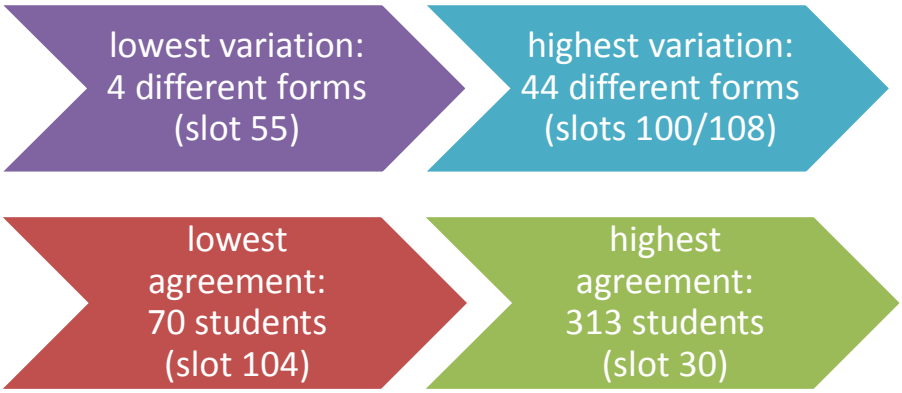


Figure 20. The extreme ends of variation, above from the perspective of the number of different forms provided and below from the perspective of the number of students providing an identical form per slot

In contrast, the highest number of variant forms was 44, suggested to both slot 100 (**She ___ (not/even/go) into the study**) and slot 108 (**She ___ (still/shout at) by her father at nine-fifteen**). There were various spelling variants[63] in addition to unconventional tense suggestions, and there were both contracted (e.g. *didn't*) and uncontracted (e.g. *did not*) forms. Numerous versions of word order also increased variation in these slots. Thus, the number of different verb forms given per slot varied between 4 and 44.

The degree of agreement was counted from the highest number of students who gave an identical form per slot. The greatest number of students giving exactly the same form regarded slot 30 (**I ___ (tell) her**), with 313 (out of 319) students suggesting that *told* was the required form. This, then, was the slot with the greatest degree of agreement, from the perspective of the number of students agreeing on the same form. The smallest number of students giving the same form per slot was in slot 104 (**Anyway, Dorothy Stern told her sister she ___ (leave) her husband**), where only 70 students suggested *was leaving* and where there were 28 other forms offered, with several students supporting many of them. This means that the smallest degree of agreement from the perspective of the number of students agreeing on the same form was 70, and thus, student agreement per slot varied from 70 to 313.

There were 19 slots where more than 300 students answered in an identical way, and all of these were the expected correct answers. Of these 19 slots, the answer was the affirmative simple past in 18 cases, and in the remaining case (slot 81), the expected answer was the *–ing* form

[63] As Crystal (2003, 272-275) reminds us, many native speakers also have difficulty in spelling in English.

(gerund) after a preposition. This means that 94% of the students knew the correct answer to 18% of the slots. If we extend the range to 90% of the students, which is 287, we find a total of 31 slots with an identical answer, thus increasing the number of correctly answered slots to 29% (see Table 9). This extension covers another gerund, although now used after a main verb in verb complementation (slot 19), and 11 affirmative simple past forms. Thus, it seems that the students included in the study mastered the use and formation of the affirmative simple past tense the best[64]. It is also likely that such forms are the ones that students are exposed to the most, and many of the slots in the test require the simple past form as the expected answer.

Table 9. Slots where more than 90% of students gave the same answer

Slot number	**Most frequent verb form given**	**Number of students giving this form**	**Teacher evaluation[65]**
9	asked	306	1 throughout
11	told	312	1 throughout
12	called	312	1 throughout
13	noticed	306	12 times 1, one ?
15	answered	307	1 throughout
18	took	299	1 throughout
19	shouting	296	12 times 1, one 4
28	phoned	300	1 throughout
29	talked	289	11 times 1, two 2
30	told	313	1 throughout
38	decided	305	1 throughout
44	went	292	1 throughout
45	saw	303	1 throughout
47	saw	308	12 times 1, one ?
55	walked	312	1 throughout
56	met	306	1 throughout
57	reached	297	1 throughout
59	called	294	1 throughout
63	let	300	1 throughout
64	seemed	299	1 throughout
65	showed	300	1 throughout
68	stopped	293	1 throughout
69	went	308	1 throughout
80	left	289	1 throughout
81	seeing	300	1 throughout
86	bought	302	1 throughout
89	went	298	1 throughout
92	met	307	1 throughout
94	seemed	298	1 throughout
97	went	297	1 throughout
98	found	300	1 throughout

[64] Note that although the students are also highly skilled in forming the negative, the fact that some students use the contracted and others the uncontracted form means that these constitute two different verb forms, and thus the number of students per form remains much lower than 287.

[65] Here, as well as in all of the other tables, a question mark indicates either a lack of response or an unclear response. Unless otherwise indicated, these are treated similarly to questionable and inappropriate responses.

The teachers were unanimous with 27 of these cases and agreed that the form given was the best choice for the slot (but keep in mind that teachers were allowed to indicate several 'best' choices). If we exclude the two cases where a teacher had not responded at all, we are left with only two cases where there was any disagreement at all. In slot 19 (**Then Trevor stopped ___ (shout)**), one teacher thought the form suggested as the best by everyone else was entirely inappropriate, while in slot 29 (**and we ___ (talk) for ages**), two teachers thought the suggested form was acceptable, but not the best. The key in the book that the test is taken from is in complete agreement with the student answers, and although it occasionally lists several answers, it does not do so in any of these cases.

If we change the analysis criteria so that it includes 80% of the students (i.e. 255 students) who answered in an identical way, we gain 14 more slots, thus reaching a total of 45 slots (42% of all the slots). This includes 11 more instances of the affirmative simple past, one instance of the affirmative simple present, and two infinitives with *to* (Table 10). Indeed, the affirmative past simple dominates the list, but it was also the most frequently expected answer (for the expected answers, see Section 7.1.2), and we can already conclude that the students had good command over the simple past, while other forms were more challenging. Here, teacher evaluation already showed some more variation, although the teachers were still unanimous with 9 of these instances. Some of these instances allow for multiple interpretations and are discussed in greater detail in later sections.

Table 10. Slots where more than 80% but fewer than 90% of students gave the same answer

Slot number	Most frequent verb form given	Number of students giving this form	Teacher evaluation
7	wanted	265	12 times 1, one 4
23	heard	283	1 throughout
24	came	278	12 times 1, one 2
26	heard	272	11 times 1, one 2, one 4
34	spilled	260	12 times 1, one 2
43	took	286	1 throughout
52	remembered	282	1 throughout
71	got	283	1 throughout
73	to have	260	1 throughout
78	realised	283	1 throughout
88	went	282	10 times 1, one 2, two 4
90	lost	273	1 throughout
102	means	271	1 throughout
105	to murder	267	1 throughout

7.1.2 The book key as the norm

If the source book answers are used as the basis for evaluation criteria, half of the slots require the simple past tense as the correct answer. Of these, 80% are affirmative forms, while one fifth has the negative simple past with *didn't* in all but three instances (the book key systematically suggests contracted forms for the negative). In some of these slots, an alternative form is indicated as well, for example one with the progressive or one including a modal auxiliary (see Table 11). Overall, the past tense prevails in the task, with 65% of the slots requiring either the simple or the progressive past tense and 10% requiring the past perfect. The present tense is expected in 6% of the cases, while only 3% require the present perfect. A modal auxiliary is allowed in 5% of the slots, and the remaining 12% consists of gerunds, infinitives with or without *to*, and structures such as *used to* and *be going to*.

Table 11. Form distribution of the suggested correct answers in the source book

Suggested form	Total (%)	Slot numbers
affirmative present simple	5 (4%)	39, 40*, 42, 83, 102
affirmative present progressive	2 (2%)	40*, 109*
affirmative past simple	49 (40%)	7, 8*, 9, 11, 12, 13, 15, 18, 23, 24, 26, 28, 29, 30, 34, 37, 38, 43, 44, 45, 47, 52, 53*, 55, 56, 57, 59, 60, 63, 64, 65, 68, 69, 71, 75, 78, 80, 85*, 86, 88, 89, 90, 92, 94, 97, 98, 103* 107, 110
negative past simple	13 (10%)	5, 6, 22, 36, 49*, 51, 74*, 76, 82, 96, 99*, 100, 106*
affirmative past progressive	19 (15%)	8*, 10, 16, 17, 25, 27, 32, 35, 46, 48, 50, 54*, 58, 67*, 77, 91, 93, 104*, 108
affirmative present perfect simple	1 (1%)	84*
affirmative present perfect progressive	2 (2%)	84*, 109*
affirmative past perfect simple	10 (8%)	4, 14*, 20, 31, 41, 53*, 61, 70, 95, 101
affirmative past perfect progressive	3 (2%)	14*, 62, 67*
affirmative modal auxiliary	1 (1%)	103*
negative modal auxiliary	5 (4%)	49*, 74*, 99*, 106*, 106*[66]
'to' + infinitive	5 (4%)	21, 72, 73, 87*, 105
infinitive without 'to'	1 (1%)	33
'used to' + infinitive	1 (1%)	85*
'be going to' + infinitive	2 (2%)	54*, 104*
present participle or gerund (-ing)	5 (4%)	19, 66, 79, 81, 87*

* In this slot, the book accepts several forms.

[66] There are, indeed, two different modal structures offered for slot 106.

Table 12 lists all the forms that the source book lists as correct answers together with the number of students giving this form and the teachers' evaluation regarding the form. Three answers suggested in the source book are not given by any of the 319 students, although some students used a parallel form, such as *have been living* in 84 (**I ___ (live) here for two years now)** and *cannot have been* in 99 (**It ___ (not/be) his wife)**. The source book gives the contracted forms (*can't*, *didn't* etc.) as the default answer for slots requiring negative verb forms.

Table 12. The source book answers with the number and percentage of students giving that answer and teachers' evaluation of the form

Slot	Alternative	Number of students	%	Teacher evaluation
4	had been murdered	68	21	1111111114111
5	didn't love	200	63	1111111111111
6	didn't murder	204	64	111111111131
7	wanted	265	83	1111111411111
8	had	219	69	4131443111111
	was having	3	1	4221444411?24
9	asked	306	96	1111111111111
10	was watching	153	48	1111111111111
11	told	312	98	1111111111111
12	called	312	98	1111111111111
13	noticed	306	96	111111111?1
14	had expected	44	14	4111114111111
	had been expecting	3	1	4121141111221
15	answered	307	96	1111111111111
16	was still shouting	189	59	1111111111111
17	were obviously having	60	19	1111111111111
18	took	299	94	1111111111111
19	shouting	296	93	1114111111111
20	had gone	81	25	1211111111141
21	to go	239	75	1111111111111
22	didn't want	223	70	1111111111111
23	heard	283	89	1111111111111
24	came	278	87	1111211111111
25	was still talking	247	77	1111111111111
26	heard	272	85	1111111211141
27	wasn't shouting	96	30	2111111111112
28	phoned	300	94	1111111111111
29	talked	289	91	1111212111111
30	told	313	98	1111111111111
31	had decided	186	58	1111111111111
32	was watching	145	45	1111111111211
33	take	172	54	1111111111111
34	spilt	7	2	4414141141444
35	was pouring	170	53	1111111111111
36	didn't want	237	74	1111111111111
37	crept	117	37	1111111111111
38	decided	305	96	1111111111111
39	never like	107	34	4411224114412
40	talk	131	41	4411114114411
	am talking	63	20	4421214111131

Slot	Alternative	Number of students	%	Teacher evaluation
41	had had	58	18	4111112111121
42	normally takes	207	65	1111411111314
43	took	286	90	1111111111111
44	went	292	92	1111111111111
45	saw	303	95	1111111111111
46	was walking	90	28	2111211111214
47	saw	308	97	111111111?11
48	was standing	113	35	1111111111111
49	didn't see	214	67	2111111111111
	couldn't see	1	0	4121111113234
50	was talking	192	60	1111111111111
51	didn't answer	223	70	2111111111111
52	remembered	282	88	1111111111111
53	had told	137	43	1111111111131
	told	163	51	4122212111414
54	was playing	107	36	2121411111111
	was going to play	10	3	2112211111231
55	walked	312	98	1111111111111
56	met	306	96	1111111111111
57	reached	297	93	1111111111111
58	was looking	217	68	1111111111111
59	called	294	92	1111111111111
60	was	250	78	4111111111111
61	had planned	124	39	1111111111111
62	had been visiting	2	1	4121221111211
63	let	300	94	1111111111111
64	seemed	299	94	1111111111111
65	showed	300	94	1111111111111
66	shouting	251	79	1111111111111
67	were having	42	13	1121211141141
	had been having	0	-	-
68	stopped	293	92	1111111111111
69	went	308	97	1111111111111
70	had already left	222	70	1111111111111
71	got	283	89	1111111111111
72	to explain	230	72	1111111111111
73	to have	260	82	1111111111111
74	didn't let	213	67	2411114111332
	wouldn't let	13	4	4121211111111
75	was	229	72	1111111111111
76	didn't know	174	55	2111111111111
77	was talking	196	61	1111111111111
78	realised	283	89	1111111111111
79	arguing	101	32	4111141111111
80	left	289	91	1111111111111
81	seeing	300	94	1111111111111
82	weren't	143	45	4111111111111
83	is	239	75	1111411111111
84	've lived	2	1	4414112414111
	‘ve been living	0	-	-
85	used to have	2	1	4121111411431
	had	152	48	4111114111441
86	bought	302	95	1111111111111

Slot	Alternative	Number of students	%	Teacher evaluation
87	earning	96	30	4122112111211
	to earn	200	63	1111111111111
88	went	282	88	1111142411111
89	went	298	93	1111111111111
90	lost	273	86	1111111111111
91	was looking	173	54	1111111111131
92	met	307	96	1111111111111
93	was walking	172	54	1111111111131
94	seemed	298	93	1111111111111
95	had seen	218	68	1111111111111
96	hadn't	146	46	2114111114111
97	went	297	93	1111111111111
98	found	300	94	1111111111111
99	can't have been	0	-	-
	wasn't	127	40	2411114111341
100	didn't even go	201	63	2111111111141
101	had found out	33	10	1411112111141
102	means	271	85	1111111111111
103	left	229	72	1111111121114
	must have left	2	1	4121111121231
104	was going to leave	27	8	2111111111241
	was leaving	70	22	1121414111131
105	to murder	267	84	1111111111111
106	didn't walk	189	59	4431414244412
	can't have walked	1	0	4141114114441
	couldn't have walked	7	2	2111111111111
107	met	260	82	1111111111111
108	was still being shouted at	28	9	2111111111111
109	has been telling	5	2	4111111111231
	is telling	183	57	1141111111111
110	made	49	15	4411111411111

When using the book as the norm, the students participating in this research scored differently than when using the teachers as the norm. Now, over 90% of the students gave (one of) the suggested correct alternative(s) to 33 slots[67], while fewer than 10% of the students gave the correct answer to 4 slots. These four included *spilt* (slot 34), *had been visiting* (62), *'ve lived* (84) and *was still being shouted at* (108). However, numerous students (260 students) gave the regular past tense spelling of *spill*, i.e. *spilled*, in slot 34, and the uncontracted form of slot 84, *have lived* (254 students). Extending the negative forms to include uncontracted forms would increase the number of students who scored well; this is discussed in Section 8.5.

The teachers, however, did not fully agree with the correct alternatives listed in the book. The teachers were entirely unanimous with the book in 60 answers (50%), and only one teacher[68]

[67] In two instances (slots 53 and 87), the 90% score is the combined result of the two alternatives given.
[68] The teachers' disagreement profiles are introduced in Section 9.1.

disagreed in an additional 20 answers (17%), while only two teachers disagreed in eight more answers (7%). In the rest of the cases, the disagreement sometimes only occurred between considering the alternative acceptable but not the best, but at other times, the suggested correct alternative was considered entirely inappropriate by some of the teacher informants. In some slots, the teachers suggested another correct alternative. This is discussed further below.

7.1.3 The extent of variation in teachers' assessment

If instead of the book we use the teachers as the norm and consider correct the forms that ten or more teachers out of thirteen thought are either the best or acceptable, we gain different results (see Table 13). This contains more forms per slot than the book in several cases, but there were also three slots where there is no form that ten or more teachers would unanimously accept. These three cases were slots 39, 40 and 85. In slot 85 (**I ___ (have) a little cottage in the village**), the disagreement covered a range of forms, from *have* to *had*, *also have*, *'ve got* and *used to have*, with some even accepting *do have*. Thus, there seemed to be disagreement regarding whether Gerald still owned the cottage or whether he sold it when he bought his present house. This slot is further discussed in Sections 8.2 and 9.2.2. Slots 39 and 40 appeared in the same sentence, and in slot 39 (**I ___ (never/like) Mum or Dad to be around)**, the variation in accepted forms ranged from present to past and further to present perfect and past perfect; thus, teachers did not seem to agree on the point in time when the sentence took place. Furthermore, some teachers thought students should use the simple form in slot 40 (**when I ___ (talk) to him)**, while others recommended the progressive, but whether this was in the past or the present was not agreed on, either. Slots 39 and 40 are discussed in Sections 8.2 and 8.5 in more detail.

Table 13. The forms that at least ten teachers consider either the best or acceptable, constituting the conservative approach

Slot	Verb form	Teacher rating	Number of students	Slot total	%
4	had been murdered	1111111114111	68	68	21
5	did not love	1111111111121	77		
	didn't love	1111111111111	200	277	87
6	did not murder	1111111111111	78		
	didn't murder	1111111111131	204	282	88
7	wanted	1111111411111	265	265	83
8	was going to have	1121121111?21	15		
	was supposed to have	1121112211?21	1		
	was to have	1111113111?21	4	20	6

Slot	Verb form	Teacher rating	Number of students	Slot total	%
9	asked	1111111111111	306		
	asked me	2121212212421	1	307	96
10	was watching	1111111111111	153	153	48
11	told	1111111111111	312	312	98
12	called	1111111111111	312	312	98
13	noticed	111111111?1	306	306	96
14	expected	1114324112211	182		
	had been expecting	4121141111221	3		
	had expected	4111114111111	44	229	72
15	answered	1111111111111	307	307	96
16	was still shouting	1111111111111	189	189	59
17	were obviously having	1111111111111	60	60	19
18	took	1111111111111	299	299	94
19	shouting	1114111111111	296	296	93
20	had gone	1211111111141	81		
	went	1121234211242	164	245	77
21	to go	1111111111111	239	239	75
22	did not want	1111111111111	75		
	didn't want	1111211111432	223	298	93
23	heard	1111111111111	283	283	89
24	came	1111211111111	278	278	87
25	was still talking	1111111111111	247	247	77
26	heard	1111111211141	272	272	85
27	was not shouting	1111111114111	58		
	wasn't shouting	2111111111112	96	154	48
28	phoned	1111111111111	300		
	phoned me	1221111211421	4	304	95
29	talked	1111212111111	289	289	91
30	told	1111111111111	313	313	98
31	had decided	1111111111111	186	186	58
32	was watching	1111111111211	145	145	45
33	take	1111111111111	172	172	54
34	spilled	1111111111112	260	260	82
35	was pouring	1111111111111	170	170	53
36	did not want	1111111111111	65		
	didn't want	2111111111411	237	302	95
37	crept	1111111111111	117	117	37
38	decided	1111111111111	305	305	96
39	-	-	-	-	-
40	-	-	-	-	-
41	had had	4111112111121	58	58	18
42	normally takes	1111411111314	207		
	normally takes a	4114111111?41	5	212	66
43	took	1111111111111	286	286	90
44	went	1111111111111	292	292	92
45	saw	1111111111111	303	303	95
46	walked	1121114111131	221		
	was walking	2111211111214	90	311	97
47	saw	111111111?11	308	308	97
48	was standing	1111111111111	113	113	35
49	did not see	1111114111111	71		
	didn't see	2111111111111	214	285	89
50	was talking	1111111111111	192	192	60

Slot	Verb form	Teacher rating	Number of students	Slot total	%
51	did not answer	1111111111111	65		
	didn't answer	2111111111111	223	288	90
52	remembered	1111111111111	282		
	remembered that	1111111111121	2	284	89
53	had told	1111111111131	137		
	told	4122212111414	163	300	94
54	was going to play	2112211111231	10		
	was playing	2121411111111	107		
	would be playing	2111112111241	21	138	43
55	walked	1111111111111	312	312	98
56	met	1111111111111	306	306	96
57	reached	1111111111111	297	297	93
58	was looking	1111111111111	217	217	68
59	called	1111111111111	294	294	92
60	was	4111111111111	250	250	78
61	had planned	1111111111111	124	124	39
62	had been visiting	4121221111211	2		
	had visited	4111114121121	79	81	25
63	let	1111111111111	300	300	94
64	seemed	1111111111111	299	299	94
65	showed	1111111111111	300	300	94
66	shouting	1111111111111	251	251	79
67	had	1421114111214	149		
	had had	4111312114221	39		
	were having	1121211114141	42	230	72
68	stopped	1111111111111	293	293	92
69	went	1111111111111	308	308	97
70	had already left	1111111111111	222	222	70
71	got	1111111111111	283	283	89
72	to explain	1111111111111	230	230	72
73	to have	1111111111111	260	260	82
74	did not let	1111114111322	63		
	would not let	4121211111111	8		
	wouldn't let	4121211111111	13	84	26
75	was	1111111111111	229	229	72
76	did not know	1111111111111	57		
	didn't know	2111111111111	174	231	72
77	was talking	1111111111111	196	196	61
78	realised	1111111111111	283		
	realized	1111411411111	30	313	98
79	arguing	4111141111111	101	101	32
80	left	1111111111111	289	289	91
81	seeing	1111111111111	300	300	94
82	were not	1111111111111	113		
	weren't	4111111111111	143	246	80
83	is	1111411111111	239	239	75
84	have been living	1121111111244	27		
	have lived	1111112111111	254	281	88
85	-	-	-	-	-
86	bought	1111111111111	302	302	95
87	earning	4122112111211	96		
	to earn	1111111111111	200	296	93
88	went	1111142411111	282	282	88
89	went	1111111111111	298	298	93

Slot	Verb form	Teacher rating	Number of students	Slot total	%
90	lost	1111111111111	273	273	86
91	was looking	1111111111131	173	173	54
92	met	1111111111111	307	307	96
93	was walking	1111111111131	172	172	54
94	seemed	1111111111111	298	298	93
95	had seen	1111111111111	218	218	68
96	had not	1111112114111	79		
	hadn't	2114111114111	146	225	71
97	went	1111111111111	297	297	93
98	found	1111111111111	300	300	94
99	could not have been	2121211121124	2		
	couldn't have been	2131211121144	2	4	1
100	did not even go	1114111111141	52		
	didn't even go	2111111111141	201	253	79
101	found out	1121111111411	209		
	had found out	1411112111141	33	242	76
102	means	1111111111111	271		
	means that	4112114211124	1		
	would mean	4121122121341	11	283	88
103	left	1111111121114	229		
	must have left	4121111121231	2	231	72
104	was going to leave	2111111111241	27		
	was leaving	1121414111131	70	97	30
105	to murder	1111111111111	267	267	84
106	could not have walked	2111111111111	4		
	couldn't have walked	2111111111111	7	11	3
107	met	1111111111111	260	260	82
108	was still being shouted at	2111111111111	28	28	9
109	has been telling	4111111111231	5		
	is telling	1141111111111	183	188	59
110	made	4411111411111	49	49	15

When using the consensus of ten teachers as the norm for rating students' skills, more than 90% of the students succeeded in responding correctly to 39 slots, and there were four slots where fewer than 10% of the students gave the expected forms. Interestingly, of the four slots with below 10% success rate, only one slot was the same as when using the book as the norm (see Section 7.1.2). This was slot 108, *was still being shouted at*. The other three were slots 8 (*was going to have*, *was supposed to have* and *was to have*), 99 (*could not have been* and *couldn't have been*) and 106 (*could not have walked* and *couldn't have walked*). In these cases, the source book accepted different forms and thus the score was different, with 70%, 40% and 61% of the students, respectively. Slots 34, 62 and 84, which were correctly responded to by fewer than 10% of the students using the book as the norm, were now answered appropriately by 82%, 25% and 88% of the students, respectively.

If the norm consists of all the cases where at least one of the 13 teachers rated a particular verb form as either the best or acceptable, this, again, gives a very different picture of the students' skills. Table 14 gives these forms and their ratings, with the number of students providing such forms. Using Table 14 as the norm, with all the forms that at least one teacher considered the best or acceptable (see also Chapter 9), would mean that more than 90% of the students gave a correct answer to 79 slots and more than 80% gave a correct answer to an additional 22 slots. This means that only six slots had a success rate below 80%, and more than 70% gave a correct answer to four of them as well. The two slots where the success rate was below 70% were slots 58 and 108, where 69% and 26% of the students gave a correct answer. Thus, it is only slot 108 that actually stands out as being particularly difficult. Although at least one teacher accepted a total of seven different verb forms in this slot, all of them required using the passive, which seemed to be the crucial element in this slot that was difficult for many students. However, there were only two slots in the entire test where a passive form was expected, and this one came very late in the task; the first one was at the very beginning in slot 4. Therefore, we cannot form far-reaching conclusions about the skills levels in this particular feature of English grammar, particularly as the form expected in slot 108 (**She ___ (still/shout at) by her father at nine-fifteen)** included adding the word *still* in the right place and using the progressive past tense with a preposition at the end as well. Although the teachers also accepted answers without the progressive aspect, without the word *still* and with different tenses, only 82 students out of 319 provided a form that teachers rated to be acceptable or the best.

Table 14. Verb forms with at least one teacher rating them as the best or acceptable, constituting the lenient approach

Slot	Verb form	Teacher rating	Number of students	Slot total	%
4	had become murdered	3232434444444	1		
	had been murded	4444444441444	1		
	had been murdered	1111111114111	68		
	was murderd	4444444444424	1		
	was murdered	2424234224422	173		
	was murdured	44444?4444424	1	245	77
5	am not in love with	4444444444422	1		
	did not love	1111111111121	77		
	didn't love	1111111111111	200		
	do not love	4444434444412	3		
	don't love	4444434444412	20		
	have not loved	443423444 3?2	1		
	I did not love	4421441122124	1		
	wasn't in love with	2122244412422	1	304	95
6	did not	44?1444444444	1		
	did not murder	1111111111111	78		
	didn't kill	4231444432424	1		

Slot	Verb form	Teacher rating	Number of students	Slot total	%
6 cont.	didn't murder	1111111111131	204		
	didn't slaughter	4231444434434	1		
	have not murdered	4344214224422	3	288	90
7	wanted	1111111411111	265		
	wanted to	4434421422434	1		
	wants	4144424444444	31		
	wished	4231414434434	1		
	would want	4224314444444	2	300	94
8	had	4131443111111	219		
	had had	4444444444424	2		
	has	4444414444434	10		
	should have	4224344444424	1		
	was about to have	3122344424422	1		
	was going to have	1121121111?21	15		
	was having	4221444411?24	3		
	was supposed to have	1121112211?21	1		
	was to have	1111113111?21	4		
	would have	4124414444?44	28		
	would have had	1441224442?31	21		
	would've had	2441224442?42	1	306	96
9	asked	1111111111111	306		
	asked me	2121212212421	1		
	asks	4444424444434	2		
	did ask	2134334444442	2		
	had asked	4224244424222	2		
	was asking	4424334424424	2	315	99
10	had watched	4443324444424	3		
	was waching	4444444424444	2		
	was watching	1111111111111	153		
	watched	4434344244414	152	310	97
11	tell	4444424444444	3		
	told	1111111111111	312	315	99
12	called	1111111111111	312		
	calls	4434424244434	2		
	made a call	4423344434432	1	315	99
13	did not notice	4444444444424	1		
	did notice	2231224244442	3		
	didn't notice	4444444444414	2		
	had noticed	4424444424232	3		
	notice	4444424444444	2		
	noticed	111111111?1	306	317	99
14	expect	4444414444444	3		
	expected	1114324112211	182		
	had been expecting	4121141111221	3		
	had expected	4111114111111	44		
	was expecting	4244444444444	4		
	were expected	4444441444434	4		
	were expecting	1124244412421	65	305	96
15	answer	4444424444444	2		
	answerd	4444444424444	4		
	answered	1111111111111	307	313	98
16	is still shouting	4434424444434	2		
	shouted	4444444444424	1		
	shouted still	4234244444444	13		

Slot	Verb form	Teacher rating	Number of students	Slot total	%
16 cont.	still shouted	4224444444332	66		
	still was shouting	4133444423241	1		
	was shouting	4134444422234	3		
	was shouting still	4134434424342	2		
	was still shouting	1111111111111	189	277	87
17	had obviously	4134144444432	56		
	had obviously had	2334142224442	16		
	obviously did have	4234344244444	1		
	obviously had	3234244224432	139		
	obviously had had	4424244424432	2		
	obviously were having	4123334222141	1		
	obviousy had	4444444424444	1		
	seemed to have	4431444444444	1		
	was obviously having	4444444424444	1		
	were obviously having	1111111111111	60		
	were obviouslyt having	4444444424444	1	279	87
18	did take	4424334444444	2		
	take	4444424444444	3		
	took	1111111111111	299	304	95
19	and shouted	4444444444424	1		
	shouting	1114111111111	296		
	shoutting	4444444424444	3		
	the shouting	4121444223242	1		
	to shout	4444144444444	11	312	98
20	goes	4444424444444	1		
	had gone	1211111111141	81		
	was going	2434344224434	1		
	went	1121234211242	164	247	77
21	go	4143444444444	24		
	into going	4121214121434	6		
	to go	1111111111111	239		
	to go to	4414444422224	1	270	85
22	did not want	1111111111111	75		
	didn't want	1111211111432	223		
	didn't wont	4444444444244	1		
	didnt want	4444444444244	2		
	do not want	4444444444424	1		
	wanted not	4444444444442	1	303	95
23	did hear	2134334444442	1		
	had heard	4424444444434	1		
	hear	4434424444444	3		
	heard	1111111111111	283	288	90
24	came	1111211111111	278		
	did come	4234344444444	1		
	had come	4414111114231	27		
	was coming	4434444424234	3		
	went	4444444414424	1	310	97
25	still talked	4434?44244444	49		
	still was talking	4143344423232	4		
	was still talking	1111111111111	247		
	was talking	4244444434444	4	304	95
26	did hear	4234234444444	1		
	had heard	1424444124441	6		
	hear	4444424444444	3		

Slot	Verb form	Teacher rating	Number of students	Slot total	%
26 cont.	heard	1111111211141	272	282	88
27	did not shout	2224244444442	24		
	didn't shout	2222244444422	115		
	stopped shouting	4442342424444	2		
	was not shouthing	4444444421244	1		
	was not shoutidg	4444444444244	1		
	was not shouting	1111111114111	58		
	wasn't shouting	2111111111112	96		
	wasnt shouting	4444444424244	1	298	93
28	called	4121444422421	5		
	called me	3221444422431	2		
	had phoned	4424444444444	3		
	phoned	1111111111111	300		
	phoned me	1221111211421	4	314	98
29	hadn't talked	4444444244424	1		
	talk	4444424444444	4		
	talked	1111212111111	289		
	were talking	4233131114424	18	312	98
30	did tell	4234344444444	1		
	had told	4424444424444	1		
	tell	4444424444444	1		
	told	1111111111111	313		
	was telling	2334344242434	1	317	99
31	've decided	4434424444224	1		
	decided	4424444244424	78		
	had decided	1111111111111	186		
	have decided	4434224444444	26		
	was deciding	4444444444424	1	292	92
32	had watched	4434344444424	1		
	was watchig	4444444424444	1		
	was watching	1111111111211	145		
	watched	4233344124114	164	311	97
33	take	1111111111111	172		
	to take	4144434444444	100	272	85
34	did spill	4234344444444	1		
	had spilled	4124311114411	30		
	I had spilled	4424444424444	1		
	spilled	1111111111112	260		
	spilt	4414141141444	7		
	was spilling	4134444424444	3	302	95
35	had poured	4444444244444	1		
	poured	1124444424412	118		
	poured it	44244444244?4	1		
	was pouring	1111111111111	170	290	90
36	didn't	444?444434244	1		
	did not want	1111111111111	65		
	didn't want	2111111111411	237		
	didn't want to	4414441422414	1		
	didn't want to watch	4414441422434	1		
	didnt want	4444444424444	2	307	96
37	creaped	4444444424444	1		
	creepd	4444444424444	2		
	creeped	4444444414344	171		
	crept	1111111111111	117	291	91

Slot	Verb form	Teacher rating	Number of students	Slot total	%
38	decide	4444424444444	2		
	decided	1111111111111	305		
	had decided	1124211123431	9	316	99
39	did never like	4144344444444	2		
	don't ever like	4434444424434	1		
	don't like	4424444424434	1		
	had never liked	4444441244444	10		
	have never liked	4441114121121	82		
	haven't ever liked	4441444444444	1		
	like never	4444444441444	2		
	never have liked	4442414421231	2		
	never like	4411224114412	107		
	never liked	1134412121342	101	309	97
40	'm talking	442141441142?	3		
	am talking	4421214111131	63		
	talk	4411114114411	131		
	talked	1144224444144	59		
	was talking	1141311421244	50	306	96
41	did have	4434344444424	2		
	had	1422213214214	221		
	had had	4111112111121	58		
	were having	4434344244?4	4	285	89
42	has normally taken	4144424224434	1		
	normally takes	1111411111314	207		
	normally takes a	4114111111?41	5		
	normally took	2434444214424	33		
	normaly takes	4444444424434	3		
	takes normally	4434434421444	35		
	takes normally a	4134234421244	1		
	took normally	4434444424434	6		
	will normally take	4434444244434	1		
	would normally take	4121411221332	8	300	94
43	did take	4124424444444	2		
	took	1111111111111	286		
	took me	1311111211334	1	289	91
44	did not go	4444424444444	1		
	got	4444444424444	1		
	had gone	4424344424444	5		
	have gone	4144444444444	3		
	was going	4434444224441	1		
	went	1111111111111	292		
	wouldn't have gone	4444424444434	1	304	95
45	had seen	4424444424442	2		
	saw	1111111111111	303		
	sees	4444424444444	1	306	96
46	walked	1121114111131	221		
	was walking	2111211111214	90	311	97
47	did see	2134221241442	5		
	saw	1111111111?11	308	313	98
48	stood	4122214444312	171		
	was standing	1111111111111	113	284	89
49	couldn't see	4121111113234	1		
	didn't	4144444414444	1		
	did not see	1111114111111	71		

Slot	Verb form	Teacher rating	Number of students	Slot total	%
49 cont.	didn't see	2111111111111	214		
	doesn't see	4444424444444	1		
	had not seen	4424444444444	1		
	hadn't seen	4424444444444	1	290	91
50	talked	4232434444444	120		
	was talkin	4444444424444	1		
	was talking	1111111111111	192	313	98
51	did not answer	1111111111111	65		
	didn't answer	2111111111111	223		
	hadn't answered	2444444444444	1		
	wasn't answering	4434444224432	1		
	was not answering	4434444224432	1	291	91
52	did remember	4434224444444	2		
	remember	4444424444444	17		
	rememberd	4444444424444	4		
	remembered	1111111111111	282		
	remembered that	1111111111121	2		
	remembred	4444444424444	1	308	96
53	had told	1111111111131	137		
	told	4122212111414	163		
	told me	4422442421444	1	301	94
54	'd play	4424144444442	1		
	had played	4444444444424	1		
	had to play	4421224224144	2		
	is going to play	4134424444444	1		
	is playing	4134424444444	11		
	played	1441444444444	47		
	plays	4144424444434	17		
	should play	4424334424444	1		
	was going to play	2112211111231	10		
	was playing	2121411111111	107		
	was to play	4121314221141	1		
	woud play	4444444424444	1		
	would be playing	2111112111241	21		
	would play	1121134421141	60	281	88
55	had to walk	4221212423324	1		
	walked	1111111111111	312		
	went	4431444434444	2	315	99
56	meet	4444424444444	7		
	met	1111111111111	306	313	98
57	had reached	4434244444424	6		
	reach	4444424444444	4		
	reached	1111111111111	297		
	was reaching	4424414444444	7	314	98
58	was looking	1111111111111	217		
	was looking for	4414441421444	1		
	was lookking	4444444424444	1	219	69
59	call	4444424444444	2		
	called	1111111111111	294		
	did call	4234414441444	1		
	had a call	4444444424444	1		
	had called	4424444424434	1	299	94
60	'd be	2444444444444	1		
	am	4444424444444	2		

Slot	Verb form	Teacher rating	Number of students	Slot total	%
60 cont.	was	4111111111111	250		
	was being	4444244444444	5		
	was going to be	4434444424442	1		
	was going to be there	4434444424444	1		
	was there	4431414424434	2		
	would be	1424444444444	18		
	would have been	4424444444444	2	282	88
61	did plan	4444424444444	1		
	had planned	1111111111111	124		
	have planned	4444424444444	2		
	planned	1424222123412	136		
	was planning	4434444223434	3	266	83
62	had been visiting	4121221111211	2		
	had had to visit	4121211423224	1		
	had to visit	4431222124224	9		
	had visited	4111114121121	79		
	visited	1224224224134	188		
	was visiting	4234324214414	23	302	95
63	let	1111111111111	300		
	lets	4444424444444	2		
	was letting	4444244424444	3	305	96
64	saw	4424444444444	2		
	seemed	1111111111111	299		
	seemed to be	4311121422324	2		
	seems	4444424444444	3	306	96
65	showd	4444444424444	2		
	showed	1111111111111	300	302	95
66	shout	4124144244432	25		
	shouting	1111111111111	251		
	was shouting	4422214121442	19	295	92
67	are having	4434224444442	7		
	did have	4424224444444	1		
	had	1421114111214	149		
	had had	4111312114221	39		
	had some	4341414424244	1		
	have	4444314441444	35		
	have had	4434344244242	28		
	were having	1121211114141	42	302	95
68	stopped	1111111111111	293	293	92
69	did go	4434424444434	2		
	go	4444424444444	1		
	went	1111111111111	308	311	97
70	already had left	4133444421441	3		
	had allready left	4414444424444	1		
	had already left	1111111111111	222		
	had alredy left	4414444424444	4		
	has already left	4444424444444	11	241	76
71	get	4444424444444	16		
	got	1111111111111	283		
	went	4442444444444	17	316	99
72	explaining	4442414111312	34		
	to exlpain	4444444424444	1		
	to explain	1111111111111	230	265	83
73	have	4444244444444	29		

Slot	Verb form	Teacher rating	Number of students	Slot total	%
73 cont.	having	4424444444444	13		
	to have	1111111111111	260	302	95
74	did not let	1111114111322	63		
	did not let me	4414444421444	2		
	didn't let	2411114111332	213		
	didn't let me	4114444421444	10		
	didnt let	44?4444424444	1		
	doesn't let	4444424444444	2		
	would not let	4121211111111	8		
	would not let me	4424441421444	1		
	wouldn't let	4121211111111	13	313	98
75	am	4242414441442	61		
	be	4441444444444	1		
	had been	4444424444444	5		
	was	1111111111111	229		
	was being	4224444241444	3	299	94
76	did not know	1111111111111	57		
	didn't know	2111111111111	174		
	does not know	4122224441442	11		
	doesn't know	4122224441342	28	270	85
77	had talked	4434314444444	2		
	is talking	4132424441442	23		
	talked	4441114444444	49		
	talks	4442444444444	12		
	was talking	1111111111111	196		
	was talking about	4414441421444	5	287	90
78	had realised	4424444444444	1		
	realised	1111111111111	283		
	realized	1111411411111	30	314	98
79	argueing	4444414424444	23		
	arguing	4111141111111	101		
	to argue	1442144121131	158	282	88
80	did leave	4234424444444	1		
	left	1111111111111	289	290	91
81	seeing	1111111111111	300	300	94
82	are not	4442444441411	7		
	aren't	4444444441211	3		
	had not been	4424244444224	3		
	hadn't been	4424244444224	5		
	were not	1111111111111	113		
	weren't	4111111111111	143		
	werent	4414444444444	3		
	wern't	4444444424444	1	278	87
83	being	4444444444424	2		
	is	1111411111111	239		
	was	4442144414444	71	312	98
84	've lived	4414112414111	2		
	have been living	1121111111244	27		
	have lived	1111112111111	254	283	89
85	've got	4441414414414	2		
	also have	4441411424234	1		
	do have	4441414444442	1		
	had	4111114111441	152		
	have	1444414214112	151		

Slot	Verb form	Teacher rating	Number of students	Slot total	%
85 cont.	have had	4434444224444	3		
	used to have	4121111411431	2	312	98
86	bought	1111111111111	302	302	95
87	earning	4122112111211	96		
	to earn	1111111111111	200	296	93
88	did go	4134444444444	2		
	go	4444444414444	1		
	had gone	4424442411444	1		
	used to go	4444444424434	1		
	was going	4134411421434	6		
	went	1111142411111	282	293	92
89	goes	4441444424434	3		
	was going	4134444424444	3		
	went	1111111111111	298	304	95
90	lost	1111111111111	273	273	86
91	looked	4424444444414	138		
	was looking	1111111111131	173		
	were looking	4244444444444	3	314	98
92	meet	4444424444444	3		
	met	1111111111111	307		
	saw	4431444434444	1	311	97
93	walked	4134444444414	141		
	was walking	1111111111131	172	313	98
94	did seem	4134224443424	1		
	seemd	4444444424444	2		
	seemed	1111111111111	298		
	seemed to be	4121111421234	6		
	seems	4444424444444	3		
	seems to be	4444424444444	1	311	97
95	had seen	1111111111111	218		
	saw	4424424424434	38		
	would have seen	4444224444444	2	258	81
96	did not	4444444424424	1		
	didn't	4444424424324	6		
	had not	1111112114111	79		
	had not seen	4423444441444	1		
	had not seen it	4421112224424	1		
	hadn't	2114111114111	146		
	hadn't seen	4424244441444	1		
	hadn't seen it	4421212224444	1	236	74
97	did go	4434424444444	1		
	had gone	4424444444444	1		
	went	1111111111111	297	299	94
98	did find	4434424444444	1		
	found	1111111111111	300	301	94
99	can not be	4142414144444	6		
	can not have been	2121414124344	1		
	can't be	4121114121314	11		
	cannot be	4121114121314	3		
	could not	4434444444424	1		
	could not be	4132214121324	9		
	could not have been	2121211121124	2		
	couldn't be	4132214121324	8		
	couldn't have been	2131211121144	2		

Slot	Verb form	Teacher rating	Number of students	Slot total	%
99 cont.	is not	4144114121434	36		
	isn't	4144114121434	28		
	was not	1412114111331	56		
	wasn't	2411114111341	127		
	wouldn't be	4442414444444	1	291	91
100	did not even	44?1444444414	3		
	did not even go	1114111111141	52		
	did not go even	4442444444444	1		
	didn't even	4444444444414	1		
	didn't even go	2111111111141	201		
	didnt even go	4444444424444	2		
	doesn't even go	4444444424444	2		
	had not even gone	4321344444124	4		
	had not even got	4442444444444	1		
	hadn't even gone	4321314444124	3		
	has not even gone	4444444244434	1		
	never went	4331444433144	1		
	won't even go	4444444424444	1	273	86
101	did find out	4144214444444	1		
	finds out	4444424444444	5		
	found it out	4121212444444	16		
	found out	1121111111411	209		
	found out it	4243244444444	3		
	found that out	4131142241444	5		
	had found it out	4244444444?44	3		
	had found out	1411112111141	33		
	had found that out	4421214221444	2	277	87
102	could mean	4124414423434	1		
	does mean	4441414444444	1		
	means	1111111111111	271		
	means that	4112114211124	1		
	meant	4444444124444	11		
	would mean	4121122121341	11	296	93
103	did leave	4134214444444	3		
	had left	1421144124?4	34		
	had to leave	2441444442441	8		
	has left	4444444144434	5		
	left	1111111121114	229		
	must have left	4121111121231	2		
	should have left	4124214424444	1		
	would have left	4134214111432	1	283	89
104	'd leave	4121114444442	1		
	'll leave	4431214444444	1		
	had left	4434444414414	26		
	is going to leave	4434414444444	5		
	is leaving	4434414444444	10		
	left	4434444444414	44		
	wants to leave	44?1414444444	1		
	was going to	4411444444434	2		
	was going to leave	2111111111241	27		
	was leaving	1121414111131	70		
	was planning to leave	4121211433234	1		
	was to leave	4444414444344	1		
	will leave	4441414444434	29		

Slot	Verb form	Teacher rating	Number of students	Slot total	%
104 cont.	would leave	4121114124441	69	287	90
105	to murder	1111111111111	267	267	84
106	can't have walked	4141114114441	1		
	cannot have walked	41?1114114441	1		
	could not have walked	2111111111111	4		
	could not walk	41?1414444444	5		
	couldn't have walked	2111111111111	7		
	couln't walk	4444114444444	1		
	did not	4441444444444	3		
	did not walk	4431414244412	52		
	didn't walk	4431414244412	189		
	had not walked	1434244444444	3		
	hadn't walked	2434244444444	8		
	would not have walked	4424444444444	1		
	wouldn't have walked	4424444444444	2	277	87
107	had met	4424244244414	39		
	meets	4444424444444	1		
	met	1111111111111	260		
	saw	4441444434444	1	301	94
108	had still been shouted at	4134144444444	1		
	was being shouted at	4231444424344	1		
	was still beeing shouted at	4441444424444	2		
	was still being shouted at	2111111111111	28		
	was still getting shouted at	4141441244244	1		
	was still shouted at	1431244444444	48		
	would still be shouted at	4134444444444	1	82	26
109	did tell	4444414444444	1		
	has been telling	4111111111231	5		
	has told	4431124214212	12		
	has told us	4431124424314	1		
	is lying	4432444434444	1		
	is telling	1141111111111	183		
	tells	4444244444?4	41		
	told	1444224244424	47		
	told us	4431224444224	1		
	was telling	4431214121242	9	301	94
110	'll make	2144244444444	4		
	made	4411111411111	49		
	make	1144134141114	200		
	should make	4431334444434	2		
	will make	2144444444444	32	287	90

In the next section, I give a statistical analysis of how many errors the students make and compare students' background factors to the number of errors. I begin by using the most lenient rating, which is defined with the same criterion as above: a correct answer is an answer that at least one teacher considered either the best or acceptable, and an error is when there was unanimous consensus by the 13 respondents that a particular form was either questionable or inappropriate. Note that, for the purposes of this study, when a student has not answered a particular slot at all, this becomes interpreted as an error. The same happens if a student gives

a double answer, such as '*went*/*had gone*', or uses brackets, as in '(*had*) *left*'. This is because it is difficult to establish what the assessment would be when the student's intended answer is unclear.

Because it would not be reasonable to analyse the number of errors made under all possible assessment scenarios, for the purposes of this study, I first follow the lenient rating presented above as the norm in analysing the number of errors the students made and for comparing the score with the students' background (Section 7.2). The students' score in the lenient analysis is then, in Section 7.3, compared to what would happen if we used gradually stricter criteria, i.e. a moderate, a conservative and the strict way of counting the number of errors. Figure 21 illustrates the extent of strictness or leniency resulting from the number of teachers that were used to provide the criteria. The red inner circle illustrates the strict approach, with few acceptable answers, while the green outer circle exemplifies the most lenient approach, with many more acceptable answers. This comparison of different levels of leniency allows us to see what happens to students' proficiency ratings depending on how strict criteria are used.

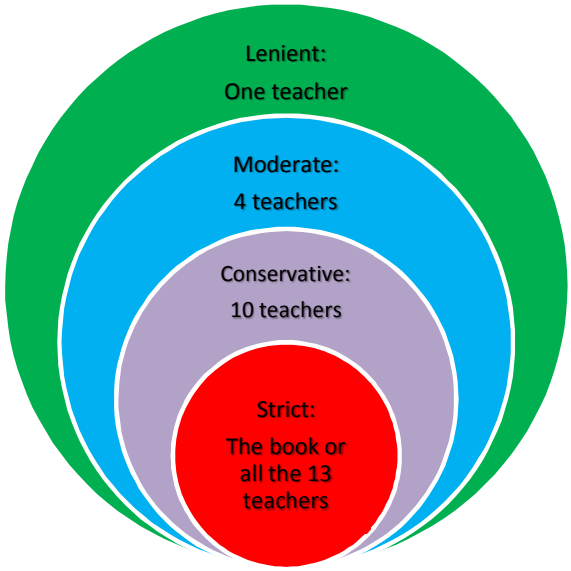


Figure 21. The extent of leniency depending on the criteria used

7.2 Student performance with the lenient approach

This section discusses the students' overall performance if rated with the most lenient approach; teachers' level of strictness is discussed in Chapter 9, and Section 9.3 addresses the different profiles teachers have regarding leniency. Figure 22 below displays the students' overall performance scores when using this lenient approach, which is the case when at least one teacher considers the form that a student has given to be either the best or acceptable. Thus, in this lenient analysis, an error is a form that all teachers unanimously consider questionable or inappropriate. As indicated in Figure 22, there were 14 students who made no errors, while the greatest number of errors was 56. Most students made 1 to 10 errors, and the mean number of errors was 8.75. More detailed analyses of the range of errors, with a discussion of factors influencing the scores, are provided below. Section 7.2.1 gives the results from analysing the entire student population and when testing for the effect of specific parameters[69]. The inter-group analysis is provided in Section 7.2.2.

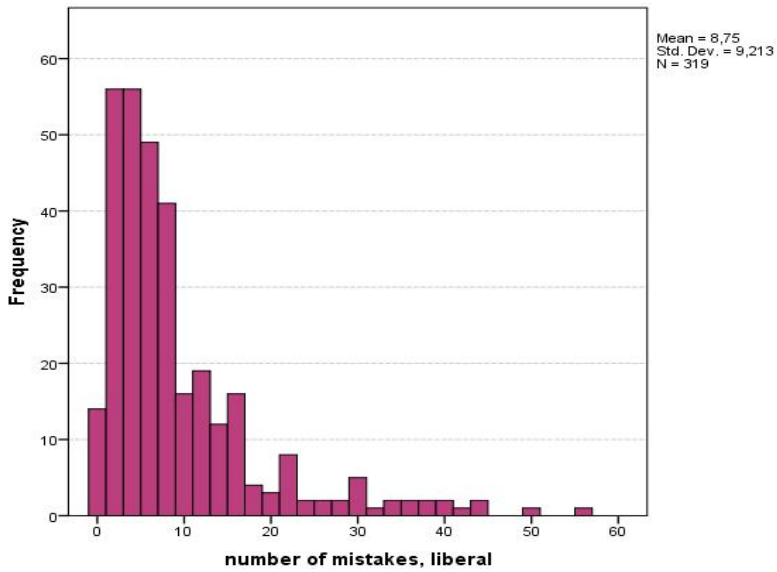


Figure 22. The number of errors in the entire student population when using the lenient approach

[69] The various statistical tests are explained in Section 6.4.3. For any post hoc analyses, the Bonferroni correction is always applied, as recommended by e.g. Kline (2004, 71) and Salkind (2008). I provide p-values, however, as this still remains the convention despite the problems associated with their use (e.g. Kline 2004, 61-91; J. D. Brown 2011), but I provide power and effect sizes when I can. The existing literature on how to do this with non-parametric data is very limited, while it is extensive for parametric data (e.g. J. D. Brown 2011).

7.2.1 Factors interacting with overall student performance

In this section, I explore the lenient results when considering the entire student population by analysing particular parameters one at a time. I examine whether particular variables seem significant from the perspective of students' skills in this test. I begin by discussing the variables that provided statistically significant effects.

The strongest link between the number of errors and background variables is the students' prior performance level. The Kruskal-Wallis test indicates that both the school-leaving mark in English and the matriculation examination mark in English are related to the number of errors the participants make ($H(6)= 153$, $p < .001$ for the matriculation examination mark and $H(6)= 134$, $p < .001$ for the school mark). This results in a strong negative correlation[70] (Spearman's rho) between the school mark and the number of errors ($r_S = -.696$, $p < .001$) as well as an even stronger negative correlation between the matriculation examination mark and the number of errors ($r_S = -.702$, $p < .001$). Since the correlations are negative, this means that the higher the mark, the fewer errors the students made. This seems to indicate that school and matriculation examination marks are a good predictor for how many errors students make in their use of verb forms in this test. In other words, the school marks provide some criterion-related validity (Gass 2015, 109; Salkind 2006; 2008) to the scores obtained in the test.

Studying languages as a major seems to be related to the number of errors in the test (Mann-Whitney, $U = 9377$, $p <.001$, $r = .274$), although to a moderate extent. None of the language students made more than 15 errors, but while some non-language students made up to 56 errors, there were also several skilful students amongst those who were not language majors. Although the command of several languages first seems to differentiate the participants' skills in English in a statistically significant way (Kruskal-Wallis, $H(6) = 25.7$, $p < .001$), the difference is statistically significant only in three out of the 21 pairwise comparisons[71], and the correlation effect is small (Spearman's rho, $r_S = -.233$, $p < .001$). Similarly, although there appears a statistically significant difference in when the students had started studying English and the number of errors (Kruskal-Wallis, $H(4) = 10.6$, $p = .032$), the difference is only statistically

[70] According to Dörnyei (2007, 223), correlations in the order of 0.6 or greater are very strong in applied linguistics research (see also Cohen, Manion and Morrison 2011).
[71] This is why the pairwise comparison statistics are not given here in full.

significant in one of the ten pairwise comparisons[72], and the correlation is minimal[73] (r_S= -.158, p = .005). Thus, these results are perhaps better explained through the influence of other (confounding) factors, such as school-leaving and matriculation examination marks or the student group (see Section 7.2.2). In addition, the students who had started their English later, in grade 4, 5 or 6, had already started another foreign language in grade 3 or earlier, which may imply that these students (or their parents) were particularly interested in languages and had perhaps invested more time in studying English. Furthermore, these students had English as their L3 and they may have benefitted from their L2 as well (see Section 2.3.1).

While some previous studies (e.g. Meriläinen 2010a; 2010b; Ringbom 1983) have found an advantage for Swedish-speaking students in learning English, in this study Swedish-speakers or bilingual students did not seem to have an advantage in the mastery of English verb forms (Kruskal-Wallis, $H(3) = 1.703$, $p = .636$). This is perhaps because of a different focus in the studies: previous studies have focused particularly on features where Finnish and English are typologically very different and Swedish and English are more similar, while this is not the case in verb forms. Furthermore, the number of Swedish-speakers and bilingual students was very small in this study. Overall, however, students whose first language was Swedish or who were bilingual tended to make fewer errors when rated with the liberal approach. The mean number of errors with the liberal approach for the Swedish-speakers and bilinguals was 6.05, while it was 8.75 for the entire population. The difference is not high enough to be statistically significant, and as we will see in Section 7.3, the advantage becomes more even when the criteria are tightened.

The students' age and the number of years that had passed since they had taken their matriculation examination have no statistically significant effect on the students' scores (Kruskal-Wallis for 'year of birth', $H(24) = 33.0$, $p = .105$ and for 'year of graduation', $H(20) = 31.2$, $p = .053$). This means that the changes in the curriculum are not reflected in students' command of verb forms, a finding similar to Meriläinen's (2010a; 2010b); however, it must be remembered that students who had graduated more recently may have benefitted from more recent exposure to English[74].

[72] This is why the pairwise comparison statistics are not given here in full.
[73] According to Pallant (2013, 129), "the practical significance of a correlation of .2 is very limited".
[74] Naturally, older students might have actively used English in their social life or at work.

Gender seems to be evenly distributed in the number of errors participants make (Mann-Whitney, $U = 12075$, $p = .480$, $r = -.040$). However, there are some differences when looking at the influence of gender at the student group level; this is illustrated in Figure 23. As we can see, men made more errors than women in the Grammar, Visual Arts and Computer Science group, while women made more errors in the Humanities group. In the English Majors and Technology groups, there was barely any difference between the genders.

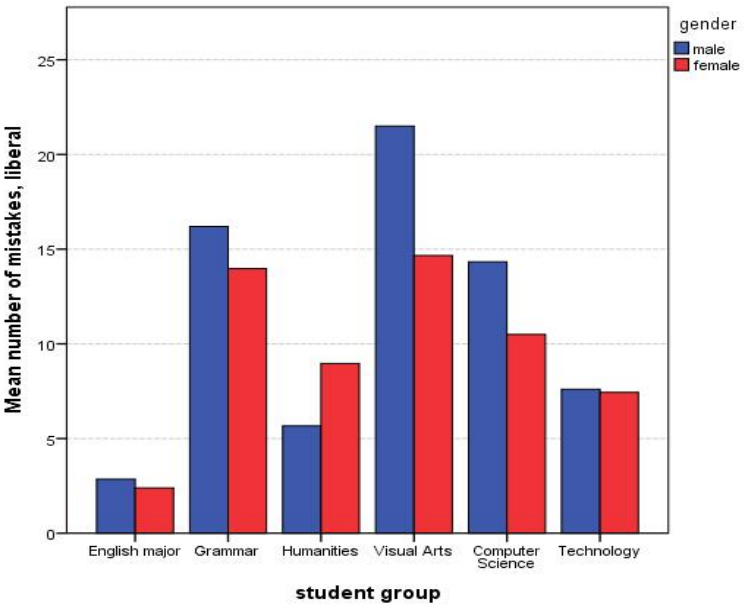


Figure 23. The mean number of errors in the various groups split according to gender

Whether the students had been engaged with English in ways other than spending time in an English-speaking country does not affect the score in the test (Kruskal-Wallis, $H(3) = .405$, $p = .939$). This seems an interesting finding, and indeed, when we explore the descriptive statistics in Table 15, we notice that although the mean number of errors is the highest for the group that used English actively, its standard deviation is the highest and its median the lowest; this means that there is notable variation within the group. Furthermore, the 'no information' and 'yes, but not language-related information' groups have high standard variation, while the passive use group is internally more coherent. Thus, at least in the light of this study, being involved in English-language activities in an active setting did not seem to influence students' accuracy in using verb forms, although it is possible that some students who were actually engaged in English-language activities did not mention their active or passive use in the questionnaire. Students who reported receptive consumption of English scored better, a finding similar to Tuokko's (2000) in Section 5.4.

Table 15. Descriptive statistics for the number of errors split along 'other information given'

	Median	Mode	Mean	Standard deviation
No information given (n=267)	6	2	8.88	9.275
Yes, passive language use (n=12)	6	several	6.00	3.303
Yes, active language use (n=20)	4.5	3	8.95	10.738
Yes, but not language-related information (n=20)	6.5	1	8.55	9.462
Total (N=319)	6	2	8.75	9.213

Another interesting feature is the fact that whether students had stayed in an English-speaking country or not did not influence the score in this study (Kruskal-Wallis, $H(18) = 24.1$, $p = .153$), although students who had stayed abroad longer seem to score better[75]. Furthermore, there is no statistically significant link between having taken additional courses in English during or after upper secondary school and the number of errors (Mann-Whitney, $U = 11715$, $p = .631$, $r = .027$). Thus, it seems that students' scores in this test were not related to the extent to which they had attempted to improve their English skills either by using the language passively or actively or by taking courses to improve their skills. Neither did extended stay in English-speaking countries seem to influence accuracy scores. Similar results have also been reported by Alanen (1997) and Bardovi-Harlig and Dörnyei (1998, 253).

Having examined the various background factors and their influence on the students' scores in the lenient approach, I now focus on the effect of the student group and discuss the extremes: the students who scored particularly well and those who were less successful. After an overview on the differences among student groups, I discuss further background factors that may influence the results as well.

7.2.2 Inter-group comparisons

Student groups seem to have given different numbers of deviant answers, and the range is wide: the standard deviation is greater than the mean for the entire student population. Table 16 lists the mean number of questionable or inappropriate answers per student in the various groups. The table is formed based on the items that all teachers rated either as questionable or

[75] Further, the number of students who had not spent time in English-speaking countries is very high.

inappropriate (the lenient approach). As we can see, students majoring in English had the lowest mean (2.49 errors per student), and they had the smallest number of errors overall. The second best group was students at Helsinki University of Technology (7.55 errors per student on average), although it is not easy to compare the groups as the number of students in them varied substantially. Perhaps surprisingly, the least successful group was not the Grammar group but the students of Visual Arts (18.08 errors per student); however, the group was also the smallest (only 12 students) and it is possible that a few low-achieving individuals distort the figures.

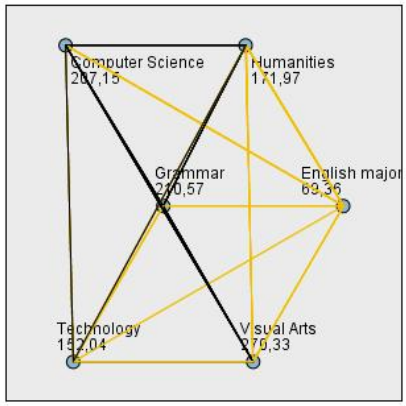
Table 16. Statistics on the errors per student group

Student group	Number of students (N)	Average of errors per student (mean)	Standard deviation (SD)	Total number of errors (sum)	Smallest number of errors (minimum)	Highest number of errors (maximum)
English Majors	35	2.49	2.15	87	0	8
Grammar	42	14.50	13.80	609	1	56
Humanities	31	8.65	8.10	268	0	43
Visual Arts	12	18.08	6.04	217	6	26
Computer Science	17	13.88	11.82	236	3	38
Technology	182	7.55	7.40	1375	0	39
Total	319	8.75	9.21	2792	0	56

The range in the number of errors was wide: while some had no errors, some students had as many as 56 responses that the teachers rated as either inappropriate or questionable, which is over 50% of the slots (although in some cases this is because the student did not provide an answer to all the slots). However, some students' language skills seemed to be considerably lower than some of their peers. The range was the narrowest within English Majors and the widest, surprisingly, within Grammar students and students in the Humanities. However, the average in the Humanities was much lower than in the Grammar group, which means that there were more low-achievers within the Grammar group than in the Humanities group. The difference between the group mean ranks was highly significant in the Kruskal-Wallis test, $H(5) = 70.3$, $p < .001$. In the post hoc pairwise comparisons, the English Majors were statistically significantly different from each of the other groups ($p < .001$ for each pairwise comparison with the English Major group). Furthermore, statistically significant differences were also found between the pairs Technology*Grammar ($p = .003$), Technology*Visual Arts ($p < .001$) and Humanities*Visual Arts ($p = .025$). The other pairs were not statistically significantly different from one another. The scores are available in Table 17.

Table 17. The pairwise comparisons with Mann-Whitney following the Kruskal-Wallis, with the student groups compared against each other group when analysing the number of errors

Pairwise Comparisons of student group



Each node shows the sample average rank of student group.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
English major-Technology	-82,687	16,981	-4,869	,000	,000
English major-Humanities	-102,611	22,691	-4,522	,000	,000
English major-Computer Science	-137,790	27,198	-5,066	,000	,000
English major-Grammar	-141,214	21,057	-6,706	,000	,000
English major-Visual Arts	-200,976	30,777	-6,530	,000	,000
Technology-Humanities	19,924	17,876	1,115	,265	1,000
Technology-Computer Science	55,103	23,333	2,362	,018	,273
Technology-Grammar	58,527	15,749	3,716	,000	,003
Technology-Visual Arts	118,289	27,420	4,314	,000	,000
Humanities-Computer Science	-35,179	27,766	-1,267	,205	1,000
Humanities-Grammar	38,604	21,785	1,772	,076	1,000
Humanities-Visual Arts	-98,366	31,280	-3,145	,002	,025
Computer Science-Grammar	3,424	26,447	,129	,897	1,000
Computer Science-Visual Arts	63,186	34,688	1,822	,069	1,000
Grammar-Visual Arts	-59,762	30,115	-1,984	,047	,708

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same.
Asymptotic significances (2-sided tests) are displayed. The significance level is ,05.

The range in the number of questionable and inappropriate answers was wide, from 0 to 56 errors. Of the 319 students, 14 students made no errors; 130 students made 1-5 errors, 88 students made 6-10 errors, 40 students made 11-15 errors, and 47 students made 16 or more errors. The errors are broken into student groups in Table 18, and Figure 24 provides a visual presentation of the number of errors per student group.

Table 18. The distribution of errors per student group

Number of errors	Majors	Grammar	Humanities	Visual Arts	Computer Science	Technology	Total
0	7	-	1	-	-	6	14
1-5	25	9	9	-	5	82	130
6-10	3	13	13	1	5	53	88
11-15	-	10	5	3	1	21	40
16-20	-	2	1	3	1	7	14
21-25	-	1	1	3	1	4	10
26-30	-	1	-	2	2	4	9
31-35	-	-	-	-	1	2	3
36-40	-	2	-	-	1	3	6
41-45	-	2	1	-	-	-	3
46-50	-	1	-	-	-	-	1
50-	-	1	-	-	-	-	1
Total	35	42	31	12	17	182	319

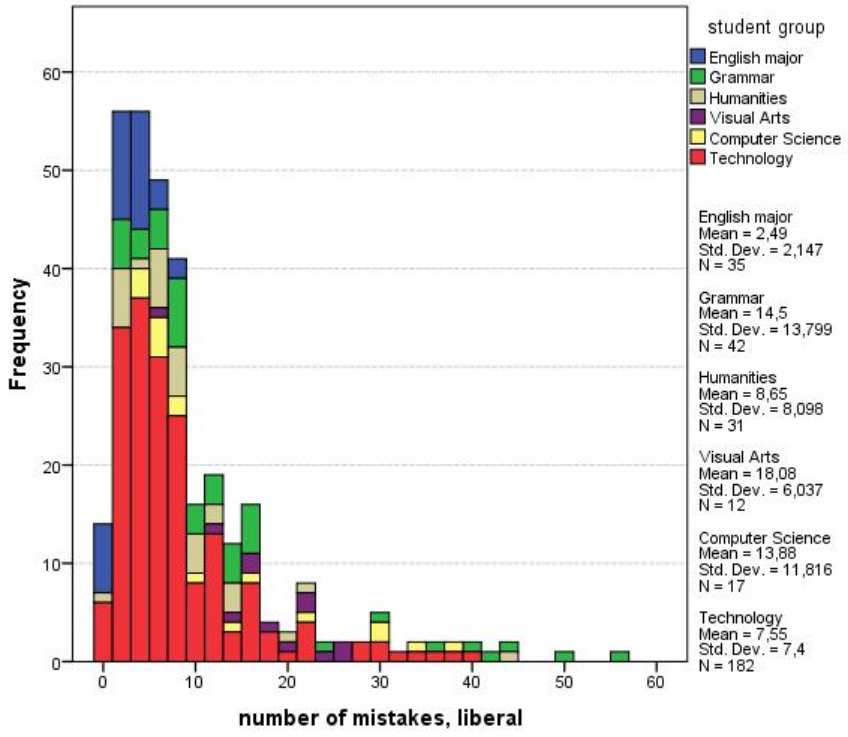


Figure 24. The cumulative frequency of the number of errors by student group

Next, the students with the highest and the lowest performance are described in more detail in relation to the background factors. Of the 14 students who made no errors, four had received the highest mark at school (10) and six the highest mark (*laudatur*) in the matriculation examination; eight students had received the second highest mark (9) at school and seven students had earned the second highest mark (*eximia cum laude approbatur*) in the matriculation examination. One student did not give her school mark, one had received the third highest school mark (8) and one student had received the third highest mark (*magna cum laude approbatur*) in the matriculation examination. Seven of these high-achieving students studied English as their major, one studied a non-language subject in the Faculty of Humanities, and six studied at Helsinki University of Technology.

At the other end, there were also 14 students who made at least 31 errors, even when rating with the most lenient norm. Of these, two students had earned mark 8 at school, four had earned mark 7, four mark 6 and two mark 5. Two students did not indicate which mark they had received. In the matriculation examination, two students had earned the third highest mark (*magna cum laude approbatur*), three had the fourth highest mark (*cum laude approbatur*), six had received the fifth highest mark (*lubenter approbatur*) and one the lowest pass mark (*approbatur*). Two students did not indicate which mark they had earned. In this low-achieving group, there were no English Majors or Visual Arts students. The low-achievers were one student from the Faculty of Humanities, two students from Computer Science, six students from the Grammar group and five students from Helsinki University of Technology.

As we have seen above, the length of time spent in English-speaking countries does not seem to correlate with the number of errors made in the test. Of the 14 students who made at least 31 errors, 11 had never been to an English-speaking country for more than short holidays. One student had stayed in an English-speaking country for one month, one student for three months and one student for five months. Similarly, of the 14 students who made no errors at all, ten had never spent even a month in an English-speaking country, while three had stayed in an English-speaking country for more than a month but less than a year (1, 3 and 9 months), and one student had spent one full year in an English-speaking country. Thus, although staying in a country where English is spoken as the native language increases the likelihood of success in English, it does not guarantee good performance in accuracy, and the fact that many students who had not stayed in an English-speaking country performed very well indicates that learners can successfully acquire a good command of English verb forms even without extensive, day-to-day contacts with native speakers. It may be that exchange periods, for example, mainly affect

students' fluency, not their accuracy. Actually, 82% of the entire respondent population (262 students) had not visited an English-speaking country for more than a month, and the percentage for the subgroup of students with more than 31 errors is 79%, which is almost identical to that of the overall population. 71% of the students with no errors had not spent much time in English-speaking countries, either. Since this is more than half of the population, we can conclude that longer visits to English-speaking countries do not in themselves increase accuracy in grammar. Note also that all of the students who had spent longer than a year in an English-speaking country made errors, ranging from 1 to 9 errors per student.

As we have seen above, there is a connection between the number of errors and school and matriculation marks in English, which is not surprising. However, the only student with a fail mark in English at school did not score too poorly: this student had 17 errors, while many students who had better marks from school scored worse. Marks 5-8 were prone to more variation in the number of errors the students make. A similar tendency is visible with the correlation between the number of errors and the matriculation examination mark, with greater variation observed with marks in the middle of the spectrum. Interestingly, however, there were two students who had the best mark in the matriculation examination yet made 18 errors, while the best two students with the lowest pass mark (*approbatur*) made 11 and 15 errors. Of course, there may be a connection here to extended visits abroad and to other means of increased use of English, but as we have seen above, extended exposure to English may not necessarily increase students' accuracy. This means that although good school marks are a good indicator of success at the group level, there are individual exceptions as well.

7.3 Student performance with the moderate, conservative and the strict approach

Section 7.2 used a lenient approach to errors to explore students' performance, and the results were based on any teacher's acceptance of particular forms. Next, I explore what happens when the criteria are tightened. First, I investigate the effects of the moderate approach, with roughly ⅓ agreement, in Section 7.3.1. Next, I move to the conservative approach, with roughly ⅔ agreement, in Section 7.3.2, and finally, the strict approach, in Section 7.3.3, aims at ultimate consensus. A comparison between the various approaches is provided in Section 7.3.4.

7.3.1 Student performance with the moderate approach

This section explores what happens to the number and nature of students' errors when we take the moderate approach. This approach is based on at least four teachers' opinion on the acceptability of the form used: the correct answers (i.e. the forms are either the best or acceptable) are now the joint opinion of not only one but at least four teachers. This means that errors are all the forms that at least nine teachers consider to be either questionable or inappropriate[76].

Table 19 presents the number of errors that students made when using this moderate approach, and Figure 25 gives a visual image of the same phenomenon. As we can see, the number of errors has increased from the lenient approach, and in some students' cases, the total number of errors is already fairly high. The mean for the entire group is 16.25. Again, Swedish-speaking students and bilingual students have a slightly lower mean, 14.81, but the difference is not statistically significant.

Table 19. The students' performance with the moderate approach

Student group	Number of students (N)	Average of errors per student (mean)	Standard deviation (SD)	Total number of errors (sum)	Smallest number of errors (minimum)	Highest number of errors (maximum)
English Majors	35	5.63	4.17	197	1	17
Grammar	42	23.12	15.68	971	3	65
Humanities	31	15.87	11.22	492	2	64
Visual Art	12	28.92	7.68	347	13	42
Computer Science	17	23.65	14.49	402	7	48
Technology	182	15.25	10.36	2776	1	58
Total	319	16.25	12.14	5185	1	65

[76] Again, failure to respond or giving a double response are considered to be errors.

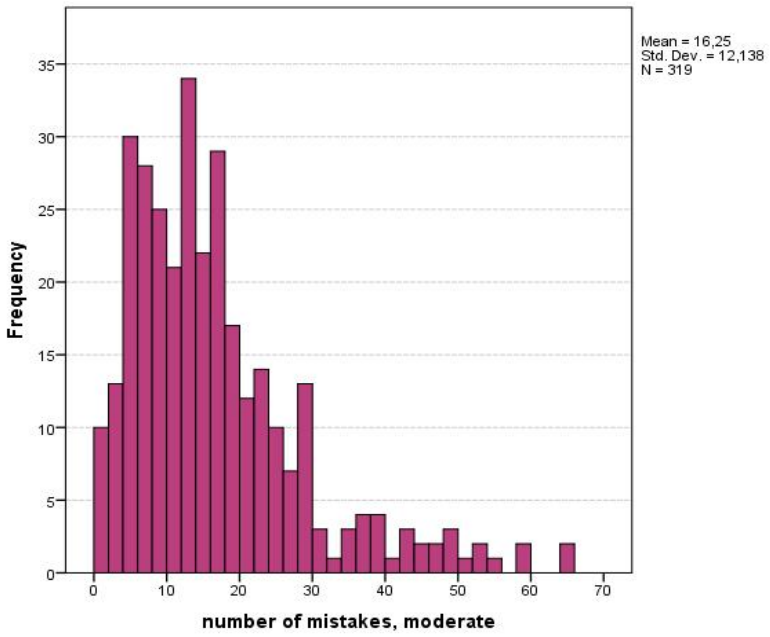


Figure 25. The distribution of the number of errors with the moderate approach

The English Majors were again noticeably different from the other groups, with a mean of 5.63 errors per student, while the mean for the other groups was well over 10. The highest mean of errors was again in the Visual Art group, with almost 29 errors per student. The Computer Science and Grammar groups had fairly similar scores, around 23 errors per student. The Humanities and Technology groups fell in the middle, with roughly 15 errors per student. Thus, tightening the criteria immediately increases the number of errors students made, and the extent is noticeable. However, while the stricter approach affects everyone, it does not treat all students in exactly the same way, since the patterns of the deviant forms they used were individual.

7.3.2 Student performance with the conservative approach

The results in the moderate approach were based on at least four teachers' approval of the forms used. This section discusses how the situation changes when we expect even more agreement: now, at least ten teachers need to agree that the form provided is either the best or acceptable for it to be rated as correct. Thus, errors are forms that at least three teachers consider either questionable or inappropriate. I call this approach the conservative approach. Now, an interesting phenomenon takes place: for three slots (slots 39, 40 and 85), no response is correct,

since there is no answer that ten teachers would consider correct (i.e. mark as the best or acceptable). Because of this, then, each student in the table below automatically "made" three errors, simply because of the lack of consensus on the correct form.

Table 20 presents the students' scores under this conservative approach, and Figure 26 shows the pattern as a graph. The increase in the number of errors was again significant, as it was from the lenient to the moderate, and even the English Majors made several errors. However, they still scored the lowest, with almost 18 errors per student, while students in the weakest group, the Visual Art group, already had 43 errors per student. Given that the total number of slots is 107, they already erred 40% of the time. The Grammar and Computer Science groups again had very similar scores, around 38 errors, and the Humanities and Technology groups resembled each other, with around 30 errors. The overall average is 30.6 errors per student, and now, the Swedish-speakers and bilinguals have the same average error number. Thus, the differences between Finnish- and Swedish-speakers as well as bilinguals were visible only at the moderate and liberal level.

Using this conservative criteria, some students already made up to 70 errors, which is ⅔ of the slots. This seems somewhat alarming, given that these students were more often wrong than right. Also, the smallest number of errors is no longer zero, but has risen to 7. Thus, even the best students in the population make errors when using the conservative criteria.

Table 20. The students' performance with the conservative approach

Student group	**Number of students (N)**	**Average of errors per student (mean)**	**Standard deviation (SD)**	**Total number of errors (sum)**	**Smallest number of errors (minimum)**	**Highest number of errors (maximum)**
English Majors	35	17.9	6.27	627	7	35
Grammar	42	37.1	14.1	1556	15	71
Humanities	31	29.9	10.9	927	14	72
Visual Art	12	43.4	8.19	521	27	57
Computer Science	17	38.3	13.6	651	21	60
Technology	182	30.2	10.5	5490	9	69
Total	319	30.6	12.2	9772	7	72

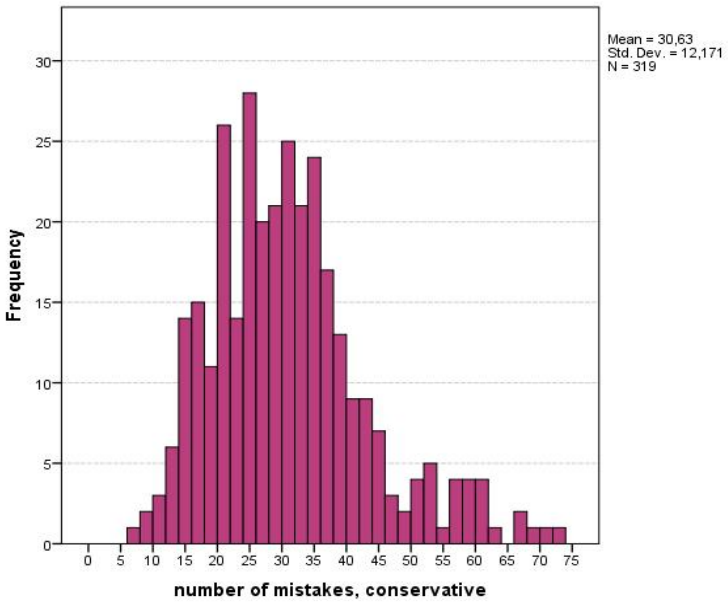


Figure 26. The distribution of the number of errors in the conservative approach

7.3.3 Student performance with the strict approach

This last approach is, in theory, the strictest: now, we only rate an answer to be correct if it is either one of the forms given in the key to the exercise or if all teachers rated it to be either the best or acceptable. As soon as one teacher marked a form questionable or inappropriate, it was dropped from the list of correct answers except if it was given in the book key. The teachers did not always agree with the book, and at times, the students never supplied a form that the book would have accepted (for examples, see Sections 7.1.2 and 7.1.3). This strict approach also holds for contracted and uncontracted forms: if a contracted form was not marked as the best or acceptable or listed in the book, it was not correct even if the corresponding uncontracted form was, and vice versa. Thus, some lack of symmetry is evident in the approved forms as a result of this strictness. The book key favours contracted forms, while teachers tended to favour uncontracted forms. Three responses that the book would allow are entirely absent since they were not offered at all. No student in the entire population used the form *had been having*, which the book would accept as an answer to slot 67 (**that they ___ (have) problems with Lucy**). The same holds for the form *'ve been living* in slot 84 (**I ___ (live) here for two years now**), although a number of students do use the form *have been living*, and for the form *can't have been* in slot 99 (**It ___ (not/be) his wife**), although some students provided *can not have been*.

As a matter of fact, for some individual students this strict approach was actually more lenient than the conservative approach, because all slots now have at least one correct answer unlike in the conservative one, where there were three slots that teachers did not agree on. Now, in this strict approach, at least the book solution always holds. Furthermore, the nature of the "correct" answers is not equal: the book, at times, recommends a response that the teachers did not approve of, and the book typically suggests contracted forms. In some instances, a critical mass of teachers approved of both contracted and uncontracted forms, but in some others, a crucial difference may cause only one of them to be accepted. However, given that the book favours contracted forms and the teachers favoured uncontracted forms, the difference does not become immense as the opposing effects are at least partly equalled out.

Table 21 presents the student data with this strict approach, and Figure 27 displays it in a graph. The English Majors persist in making the smallest number of errors, now with a mean of 19 errors per student. This is 18% of the slots. The Humanities and Technology groups again resemble one another, with around 30 errors per student, and the same holds for the Grammar and Computer Science groups, around 38 errors per student. The mean for the Visual Art group is almost 43 errors. Note that the difference to the conservative table is fairly small, with the greatest effect on the English Majors group. While the overall error rate per student is 31.3, it is now higher, 32.7, for the Swedish-speaking and bilingual students.

Table 21. The student's performance with the strict approach

Student group	**Number of students (N)**	**Average of errors per student (mean)**	**Standard deviation (SD)**	**Total number of errors (sum)**	**Smallest number of errors (minimum)**	**Highest number of errors (maximum)**
English Majors	35	19.2	7.65	671	7	36
Grammar	42	37.0	14.2	1555	12	71
Humanities	31	29.4	11.3	910	12	71
Visual Art	12	42.9	8.95	515	26	54
Computer Science	17	39.4	13.7	670	20	63
Technology	182	31.1	10.7	5658	11	67
Total	319	31.3	12.3	9979	7	71

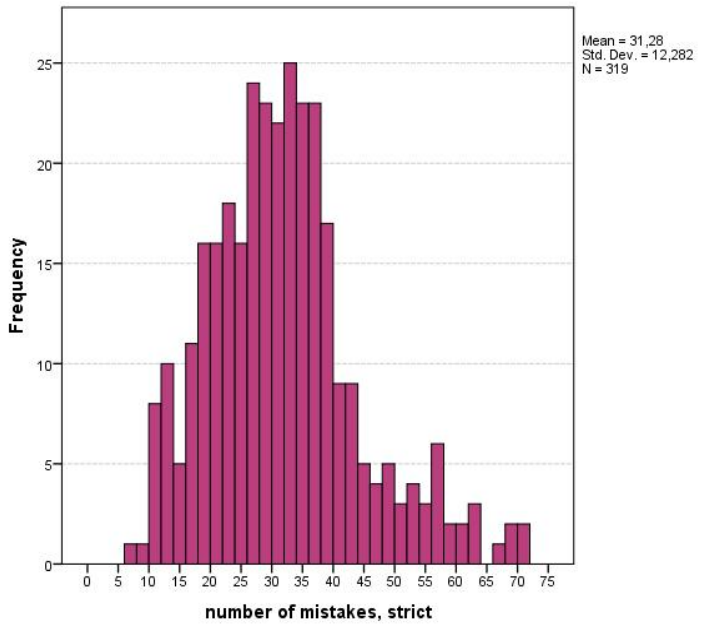


Figure 27. The distribution of the number of errors with the strict approach

Note how the shape of the graph begins to approach normal distribution, although it does not reach it (Shapiro-Wilk, $W = .963$, $p < .001$, and Kolmogorov-Smirnov, $D(319) = .085$, $p < .001$, with high values still for skewness and kurtosis). The correlation between students' school-leaving and matriculation examination marks continues to be strong across the various approaches, but it is the highest with the lenient approach and a little weaker with the strict approach (for the strict approach, the number of errors correlates with the marks with $r_S = -.575$, $p < .001$ for the matriculation examination mark and $r_S = -.605$, $p < .001$ for the school-leaving mark).

Interestingly, the difference between the English Majors and the rest of the groups becomes proportionally smaller when moving from the lenient approach towards the strict approach, although the others never close the gap. This seems to indicate that weaker students produced more forms that are never acceptable, but that students with stronger skills score well as long as some leniency is practiced; when the approach is stricter, the weaknesses in their performance becomes visible as well. This indicates that the strictness of the criteria affects the results in acceptability studies.

7.3.4 Comparison of different levels of strictness

In order to examine how the changes in strictness affect student performance in the various student groups, Table 22 was created. It illustrates the changes in the group means depending on the criteria used. As can be seen, the difference from the lenient criteria to the strict is on average 22.5 errors. This means that having a lenient teacher as the rater allows learners to make considerably many more "errors" and still to be found successful, while a stricter teacher would more readily rate their performance less favourably. The difference is very small between the conservative and the strict and more radical from the lenient to the moderate and from the moderate to either the conservative or the strict approach. The effect is somewhat weaker for the English Majors than for the rest of the groups; the English Majors clearly stand out in the comparison.

Table 22. Changes in the number of errors depending on how an error is defined

Student group	**Lenient mean**	**Moderate mean**	**Conservative mean**	**Strict mean**	**Difference from lenient to strict**[77]
English Majors (n = 35)	2.5	5.6	17.9	19.2	16.7
Grammar (n = 42)	14.5	23.1	37.1	37.0	22.5
Humanities (n = 31)	8.7	15.9	29.9	29.4	20.7
Visual Arts (n = 12)	18.1	28.9	43.4	42.9	24.8
Computer Science (n = 17)	13.9	23.7	38.3	39.4	25.5
Technology (n = 182)	7.6	15.3	30.2	31.1	23.5
Total (N = 319)	8.8	16.3	30.6	31.3	22.5

Of interest is also how students who studied a language-related major or minor score depending on the level of strictness. This is displayed in Table 23. The means for students of any language-related major/minor are higher than those for English Majors/minors, but clearly lower than those of students who do not study a language subject. While the difference between the language and non-language group is more than twofold in the lenient approach, it is proportionally more even when stricter criteria are applied.

[77] Group means are here given with only one decimal place to enable the calculation of the difference.

Table 23. The scores of language and non-language students depending on the assessment criteria

Type of major/minor	Lenient mean	Moderate mean	Conservative mean	Strict mean	Difference from lenient to strict
Language subject (n = 48)	4.0	8.4	21.6	22.6	18.6
Non-language subject (n = 271)	9.6	17.6	32.2	32.8	23.2

Another interesting comparison is illustrated in Table 24, which displays the effects of the changing criteria in the various matriculation examination mark categories. The effect of the change in strictness from the lenient to the strict was on average 22.5 errors, as we saw above. The effect of the change in strictness is noticeable, but the weakest for the highest two categories, which do not differ from each other very much. However, the overall differences between the groups are a little smaller than when comparing the changes among student groups. A similar phenomenon can be observed in Table 25, which illustrates the changes in the assessment criteria in the school-leaving marks. Note that although there is only one student in the lowest category (fail, which is marked *improbatur* in the matriculation examination and mark 4 at school) in the two tables, they are different people.

Table 24. The number of errors in the matriculation examination mark categories per different criteria

Group (N = 302)[78]	Lenient mean	Moderate mean	Conservative mean	Strict mean	Difference from lenient to strict
Laudatur (n = 46)	3.3	7.9	21.4	23.4	20.1
Eximia (n = 77)	3.7	9.3	23.9	24.5	20.8
Magna (n = 75)	7.2	14.6	29.6	29.8	22.6
Cum laude (n = 61)	11.3	20.5	35.4	35.8	24.5
Lubenter (n = 35)	18.8	29.7	43.2	43.5	24.7
Approbatur (n = 7)	26.1	38.4	50.6	50.1	24.0
Improbatur (n = 1)	30.0	39.0	52.0	52.0	22.0
Total (N = 319)	8.8	16.3	30.6	31.3	22.5

[78] There were 17 students who did not indicate their matriculation examination mark in English. They are excluded from the table proper but included in the overall total.

Table 25. The number of errors in the school mark categories per different criteria

Group (N = 274)[79]	Lenient mean	Moderate mean	Conservative mean	Strict mean	Difference from lenient to strict
Mark 10 (n = 32)	2.6	6.4	20.3	21.8	19.2
Mark 9 (n = 101)	4.2	9.8	24.4	25.0	20.8
Mark 8 (n = 79)	8.4	16.6	31.8	32.0	23.6
Mark 7 (n = 43)	15.5	25.8	40.0	40.3	24.8
Mark 6 (n = 13)	25.2	36.9	49.5	50.2	25.0
Mark 5 (n = 5)	31.2	40.4	50.8	50.6	19.4
Mark 4 (n = 1)	17.0	28.0	41.0	38.0	21.0
Total (N = 319)	8.8	16.3	30.6	31.3	22.5

Thus, it seems that school and matriculation examination marks are a good reference point for explaining how many errors in verb forms the students are likely to make. Furthermore, the students who benefit more from a lenient approach are the high-achieving students, because their scores become proportionally better than students who score worse with stricter criteria. This seems to indicate that the high-achieving students are more familiar with alternative, yet acceptable ways of expressing a verb form, while less successful students know fewer forms. Furthermore, students with good skills may be more willing to experiment with less frequent forms (see Section 8.4).

7.4 Summary

Students' overall proficiency levels depend on how the criteria for acceptable answers and errors are defined. In this study, the lenient approach means that at least one teacher finds a particular form either the best or acceptable, and an error is thus a form that all of the 13 teachers find either questionable or inappropriate. The moderate approach means that four teachers find the form the best or acceptable, while the conservative approach requires ten teachers to approve of the form. The strict approach is either the book key form or the consensus of all the 13 teachers that the form provided is either the best or acceptable. The teachers did not always

[79] There were 45 students who did not indicate their school-leaving mark in English. They are excluded from the table proper but included in the overall total.

agree with the book or each other on which forms are the best for the specific contexts, and in some slots, none of the forms gains overwhelming support.

The range in the students' responses was wide. The number of different forms provided per slot ranged from 4 to 44, and the degree of variation, measured in the highest number of students who provided the same form, varied from 70 to 313. The students were the most successful in providing the affirmative past simple form, because spelling errors as well as the use of contracted and uncontracted forms increased the number of errors with negative past simple forms. Most students were successful with both regular and irregular verbs.

Students' success rate varied depending on which criteria were applied, but the tendency remained the same. With the lenient approach, the mean number of errors was 8.75, while it was 31.3 with the strict approach. Unsurprisingly, the English Majors made the least errors. The Technology and Humanities groups competed for the second place, while the Grammar and Computer Science groups followed them. The least successful group was the Visual Arts group. In addition to the student group, the participants' marks from school and the matriculation examination correlated with test success.

In addition to the student group, the students' accuracy scores correlated the best with the students' previous test results, which may also have been based on accuracy rates. Students' extended visits, further study or active use or consumption of English did not influence the students' success in this test. Perhaps such exposure increases these students' fluency in English, but this remains to be studied further.

8 Variation in students' responses

After the quantitative analysis of students' overall proficiency levels, I delve into various phenomena arising from the responses. There are some slots that are particularly interesting and are discussed in more detail in this chapter. These forms were chosen because they emerged from the analysis and could not have been predicted beforehand. In this chapter, I am more interested in the nature of variation. However, I still need to distinguish between standard usage, the expected forms, unconventional forms and errors. The criteria for this come from the teachers' assessment, with different levels of consensus. Some slots are discussed several times from different perspectives because they are particularly susceptible to variation and evoke different responses.

First, the students' command of tense, aspect and reported speech is explored in Section 8.1. Next, some interesting phenomena arising from the study are discussed: variation in interpreting event time (Section 8.2) and variation with modal verbs, irregular verbs, the passive and verb complementation (Section 8.3). After these phenomena, I analyse the provision of unusual forms (Section 8.4) and explore whether students are systematic in their choices (Section 8.5) and discuss other, miscellaneous variation phenomena (Section 8.6). Finally, I examine how students who major in English score compared to other student groups (Section 8.7). Throughout this chapter, it is again useful to refer to Appendix 2, where the test is available. However, wherever practical, the immediate slot context is provided[80]. The full list of the responses is provided in Appendix 6.

8.1 Tense, aspect and reported speech

This section discusses variation in three verb-related phenomena: tense, aspect and reported speech. As discussed in the previous sections, it is often difficult to define what constitutes an error (see also Gass and Selinker 2001), and as we have seen, different criteria produce different

[80] I have, again, systematically dropped the slot numbers from the slot quotations.

outcomes. However, for the purposes of this discussion, I draw on what grammar books typically say about the use of verb forms in specific contexts and what they, in addition to dictionaries, say about the typical patterning of verb forms. Whether this reflects the view of the teachers in the study or the solutions in the book depends on the level of strictness or leniency you wish to adopt as well as on how extensively you are willing to consider other possibilities and interpretations, i.e. whether you tend to seek alternative ways to understand the context and whether you are sympathetic to "far-fetched" ideas as well. In addition to the assumed "standard" forms, I use the teachers' assessment of the forms students provided, as outlined above.

The discussion in this section begins with the use of the simple past in Section 8.1.1. Next, I discuss the variation between simple and progressive forms (Section 8.1.2), after which I explore the difference between the past and the past perfect (Section 8.1.3). Section 8.1.4 addresses the use of the present perfect, and finally, I investigate students' command of reported speech in Section 8.1.5.

8.1.1 Past simple forms

As we saw in Section 7.1, what seemed to be the easiest forms for students to correctly produce were simple past forms, particularly of regular verbs but also of many irregular verbs. However, some students seem to have had difficulty with irregular verb forms (see Section 8.3.2), including such common words as *hear* and *go*, where a number of students consider *hear* a regular verb and some did not know (or did not remember) that the past participle of *go* is *gone*, not *went*. Nevertheless, the majority of students did well with verb inflection. This is perhaps because many irregular verbs are in wide use, and at school, pupils are often expected to learn them by heart. Furthermore, students may be most exposed to using the past tense, since several genres that they practice reading and writing are written with the past simple as the dominant verb form. The book key expected a past simple form in 50% of the slots (see Section 7.1.2).

Table 26 provides information about the distribution of the forms given by the students in slots where the expected correct answer is the simple past tense in the active. The list is organised so that regular affirmative forms are given first, followed by verbs with both a regular and an irregular form, after which the irregular affirmative forms are given. The negative forms begin with the words *were* and *had*, after which the forms including *did* are given. Passive forms are

discussed in Section 8.3.3, and progressive forms and the past perfect are discussed below in Sections 8.1.2 and 8.1.3. The expectation of the past form is taken from the greatest consensus within teachers and the book key. Forms with more extensive teacher disagreement, which often match more extensive student disagreement, are not included in this table because they are discussed elsewhere in this study (e.g. in Sections 8.2 and 9.2).

Table 26. Slots where the simple past is the expected form (the spelling must match the expected form), excluding forms with greater teacher disagreement

Slot	Expected form	Number of students	%
7	wanted	265	83
9	asked	306	96
12	called	312	98
13	noticed	306	96
15	answered	307	96
28	phoned	300	94
29	talked	289	91
38	decided	305	96
52	remembered	282	88
55	walked	312	98
57	reached	297	93
59	called	294	92
64	seemed	299	94
68	stopped	293	92
78	realised / realized	313	98
94	seemed	298	93
34	spilled / spilt	267	84
65	showed	300	94
11	told	312	98
18	took	299	94
23	heard	283	89
24	came	278	87
26	heard	272	85
30	told	313	98
37	crept	117	37
43	took	286	90
44	went	292	92
45	saw	303	95
47	saw	308	97
56	met	306	96

Slot	Expected form	Number of students	%
60	was	250	78
63	let	300	94
69	went	308	97
71	got	283	89
75	was	229	72
80	left	289	91
86	bought	302	95
88	went	282	88
89	went	298	93
90	lost	273	86
92	met	307	96
97	went	297	93
98	found	300	94
107	met	260	82
82	were not / weren't	256	80
96	had not / hadn't[81]	225	71
5	did not love / didn't love	277	86
6	did not murder / didn't murder	282	88
22	did not want / didn't want	298	93
36	did not want / didn't want	302	95
49	did not see / didn't see	285	89
51	did not answer / didn't answer	288	90
76	did not know / didn't know	231	72
100	did not even go / didn't even go	253	79

[81] Technically, this is the auxiliary in a past perfect form where the main verb is not given at all.

The command of the use of simple past affirmative verbs was high for regular verbs, from 83% with *wanted* to 98% with *realised* / *realized*, with an average of 94% (299 students). The slightly lower percentage for *wanted* in slot 7 (**After dinner last night he said he ___ (want) to check some business papers in his study**) can be explained by the fact that 31 students used the present tense, perhaps because they either did not realise the fact that this slot included reported speech or they did not know how to mark reported speech in the verb form (see Section 8.1.5). This also affected the word *remembered* in slot 52 (**Then I ___ (remember)**), with 17 students using the present tense. Generally, however, the past simple with regular verbs posed few problems. The two words that have both regular and irregular forms (*spilled/spilt* and *showed*) were also correctly provided to a great extent, although in the past form, of course, the word *showed* has no irregular counterpart. The form *spilt* was very rare, given by only seven students. In this slot (slot 34, **He shouted at me because I ___ (spill) a few drops of tea on his desk**), the lower percentage is explained by 30 students using the past perfect form instead of the past simple. This slot is further discussed in Section 9.2.2.

The irregular past affirmative forms seemed to be in good command as well, apart from the word *crept*, which is not equally common in use as most of the other verbs. The accuracy with *told* was very high, 98%, and the average was 89% (283 students) if we include *crept* and 91% (289 students) if we do not. The lower figures for *was* in slot 60 are because 17 students used *would be* and 11 students used *were*, and in slot 75, 61 students used *am*, perhaps not knowing how to use verb tenses in reported speech (see Section 8.1.5). The Swedish-speakers showed no advantage over Finnish-speakers in the command of irregular verbs, despite the fact that in Swedish, the corresponding verbs are often irregular.

The negative forms posed some more problems, but the overall figures remained high. The students were better at using the negatives with *did* than with *had* and *were*, although the form *had not* / *hadn't* is actually the auxiliary of the past perfect, when the main verb is not produced at all in the context. This seems to have made some students confused, since 33 students provided either *didn't have* or *didn't had*. Such forms were otherwise not prevalent in the responses. The forms with *did not / didn't* were used quite confidently, with 87% success (277 students), although unconventional spelling made some students "fail" in these forms. The students were particularly successful with *did not want / didn't want* in slots 22 and 36. The lower rates for *did not know / didn't know* in slot 76 (**... who ___ (not/know) what ...**) are perhaps explained by the fact that this was again an instance of reported speech, and 33 students produced *don't know*, 28 students *doesn't know* and 11 students *does not know*.

In this study, students were not likely to provide progressive forms when they should use the simple. Of the slots in the table above, students never suggested a progressive (counted from any instance of the *–ing* form) in 26 of the slots, and only 1 to 8 students per slot did so in the remaining 27 slots, except for slot 29 (**... and we ___ (talk) for ages)**, where 23 students used a form with *-ing*. Usually, there was only one instance per student when they used the progressive where the simple would be more standard, but a number of students used this strategy more often; for example, Student 280 did this six times and Student 131 four times. In total, the progressive was used in slots where it was not expected 120 times in 40 slots[82]; this is 1% of all the forms given in those slots. In addition to slot 29, slot 106 (**Lucy ___ (not/walk) to the village and back, if ...**) was slightly more productive with the *–ing*: 13 students provided *was not walking*, *wasn't walking* or *weren't walking*. Slot 29 (... **and we ___ (talk) for ages**) was followed with '*for ages*', which possibly made some students search for the progressive and made some teachers find this a suitable response. Incidentally, the British teachers found *were talking* the best form in slot 29, while the other teachers disagreed. Perhaps the talking could also be understood as happening parallel with the next sentence (slot 30: **I ___ (tell) her I ...)**. Slot 106, however, was a less common context for the progressive, since it seemed to emphasise the result of events, not their progression: Sergeant Ross seemed to question whether Lucy walked to the village and back or whether she went there at all.

8.1.2 Past progressive forms

After the discussion of the past simple, it is time to discuss the past progressive. The book key (see Section 7.1.2) expected a past progressive form in 15% of the slots, and Table 27 displays the slots where the past progressive was the expected form, with the number of students who provided the expected form. The table also includes any other forms that seemed to be attempts at forming a progressive, and the corresponding past simple (only in the standard past simple form: e.g., for *was watching*, the standard past simple is *watched*). Slot 108 expects the passive past progressive and is discussed in Section 8.3.3. As we can see, students provided the expected past progressive form with varying success, from only 19% to 77%; the mean was 53% (168 students). This implies that roughly half of the students could provide the past progressive in contexts where it was expected. The highest proportion of students formed the past progressive in slot 25 (... **while the doctor ___ (still/talk) to Trevor**) with *was still talking*,

[82] This count covers all slots in the test as opposed to only those requiring the past tense at the beginning of this paragraph.

but the word *still* did not have a similar effect in slot 16 (**Trevor ___ (still/shout) in his study**), although the percentage there was also higher than the average. The form with *obviously* in slot 17 (**He and Lucy ___ (obviously/have) a serious row**) had the smallest proportion, only 19%. Students did not seem to readily understand the simultaneity of action between slots 16 and 17, perhaps because they were in different sentences.

Table 27. Slots where the past progressive is the expected form, excluding forms with greater teacher disagreement

Slot	Expected form	The past progressive given as expected[83]	% of all forms	Any other progressive (*-ing*) given[84]	Corresponding past simple
10	was watching	153	48%	2	152
16	was still shouting	189	59%	18	66
17	were obviously having	60	19%	8	139
25	was still talking	247	77%	19	49
27	was not shouting / wasn't shouting	154	48%	9	115
32	was watching	145	45%	5	164
35	was pouring	170	53%	10	118
48	was standing	113	35%	1	171
50	was talking	192	60%	1	120
58	was looking	217	68%	8	92
77	was talking	196	61%	48	49
91	was looking	173	54%	4	138
93	was walking	172	54%	1	141

It is unclear why students felt more tempted to provide the progressive in slot 25 (**... while the doctor ___ (still/talk) to Trevor**) than in other slots, but one possible explanation is the word *while* before the slot, which may have helped students identify a typical context for the progressive; it could also be the combined effect of *while* and *still*, since the word *while* before slot 35 (**while I ___ (pour) it**) did not have quite the same effect on the use of the progressive. The word *when* after slot 91 (**I ___ (look) for him when ...**) may also have served as a clue for using the progressive. Slot 77 (**... what he ___ (talk) about**) again shows evidence of hesitancy with reported speech, as there were a number of present progressives suggested for this slot (*is talking* by 23 students and some other forms), and the pure *–ing* form (*talking*) was given by 15 students.

[83] Double forms are not included in cases where they contain both the progressive and the simple form. When both of the forms contain the progressive, they are included.

[84] As indications of a progressive, I have here included forms ending with *–id*, *-in*, *-ig*, *–ng*, *-idg* and *–ings* in addition to *–ing*.

Slot 46 (**when I ___ (walk) into the village**) is not included in the table above, because in this slot, teachers seemed equally willing to accept the progressive as well as the simple past form, albeit with some individual variation in which was actually considered the better form. Students greatly favoured *walked* in this context, as it was given by 221 students as opposed to 90 students (28%) providing *was walking*. Other *–ing* forms were provided three times, and there were two double forms as well, with both *walked* and *was walking* offered. Perhaps these two students agreed with the teachers in finding the two equally suitable in the context.

Students whose first language was Swedish or who were bilingual scored slightly worse than Finnish-speakers with the past progressive. Their success rate varied from 19% to 62%, while the range was from 19% to 77% for the entire student population. The bilingual and Swedish-speaking students struggled particularly with slot 17 and scored the best with slots 25 and 58. Their overall accuracy rate with the past progressive remained 41%, which is lower than the mean of the entire population, which was 53%. It may be that Swedish-speakers have more difficulty associating the progressive with the structure *hålla på att göra* than Finnish-speakers with the structures *–massa* and *–essa* (see Section 5.1). However, there were only five bilingual and 16 Swedish-speaking students, and more research is needed to see if the finding depends on the L1 and whether it is generalisable.

In conclusion, students were not as proficient with the past progressive as they were with the past simple, but there was variation among the slots, and further research would be needed to explore why students preferred to provide the progressive in some of these slots. The results do not provide support for the assumption that students would be more inclined to provide the progressive with particular types of verbs (see Section 4.1). For example, while both *stand* and *talk* are activity verbs, the success rate for *stand* was 35%, but it varied from 60% to 77% for the three instances of *talk*. However, since this study did not specifically target the difference between various types of verbs and the provision of a particular aspect with them, it does not contradict the aspect hypothesis or discourse hypothesis (see Section 4.1), either.

8.1.3 Past perfect forms

Another topic to investigate in this section is the prevalence of the past perfect in contexts where the teachers and the book key expected it. Table 28 shows the proportion of using the past perfect in slots where it was the expected form. The table also includes other forms that seemed

to be attempts at forming a past perfect form, the corresponding past simple (also in the expected form, i.e. in the standard past simple form of the word given with no spelling errors, as e.g. *went* for *had gone*), and attempts at providing any present perfect form of the same verb. The book key (see Section 7.1.2) expects the past perfect simple in 8% of the forms and the past perfect progressive in 2% of the slots. Slot 4, where the expected form is a past perfect but in the passive form, is discussed in Section 8.3.3.

Table 28. Slots where the past perfect is the expected form, excluding forms with greater teacher disagreement

Slot	Expected form	The past perfect as expected[85]	%	Other past perfect attempts[86]	Corresponding past simple	Present perfect attempts[87]
20	had gone	81	25%	12	164	3
31	had decided	186	58%	10	78	32
41	had had	58	18%	6	221	19
53	had told	137	43%	2	163	6
61	had planned	124	39%	17	136	2
62	had visited / had been visiting	81	25%	5	211	0
70	had already left	222	70%	38	13	24
95	had seen	218	68%	25	38	23

Again, the percentages vary greatly: the highest proportion of the past perfect, 70%, was with the word *already* in slot 70 and the lowest, 18%, in slot 41. The mean of the past perfect was 43% (138 students). It is possible that in slot 70 (**Lucy ___ (already/leave) the room before ...**), students were prompted by the word *already* to assume that the past perfect would be the suitable form; the word *before* after the slot may have reinforced the effect. Unfortunately, this was the only instance of the word *already* in the test, and its effects cannot be explored further. In slot 62 (**... because I ___ (visit) another patient)**, both the simple and the progressive were used, although only two students provided the progressive form. The teachers accepted both forms but perhaps found the progressive marginally inferior to the simple; however, the difference was small.

[85] Double forms are excluded.

[86] This includes other forms where there is first either *had* or *hade*, followed by another verb or other verbs but not with *to* (i.e., not forms such as *had to visit*) except for *had had to visit*, which is accepted because it contains a past perfect.

[87] This includes forms where the first word is either *has*, *have* or *'ve* and which are followed by another verb or other verbs but not with *to* (i.e., not forms such as *has to visit*).

It is surprising that students did not attend to the time reference *the day before* after slot 41 (**Especially yesterday, because Dad and I ___ (have) a stupid argument about Alan the day before**) more extensively to aim at a past perfect form. Perhaps the fact that there were intervening words confused the students, or perhaps they were not fully aware of the appropriateness of the past perfect in this type of situation. However, they were much better at finding the reference to prior past time in slot 95 (**I asked her if she ___ (see) the dog)** without any lexical help. Thus, it remains unclear what caused the differences in the proportion in the provision of the past perfect, and why students tended to provide the past perfect in some of these slots but not in others. At times, students offered the present perfect, and there may, indeed, have been some confusion about the difference between the present perfect and the past perfect. Of the slots requiring the past perfect, slot 62 was the only slot where no student offered the present perfect; in some other slots, up to 10% of the students suggested the present perfect.

Two interesting cases remain to be explored: slots 14 and 101. In slot 101 (**And perhaps Trevor Stern ___ (find out)**), teachers accepted *found out* (except for one teacher) as well as *had found out* (except for two teachers). Thus, both the past simple and the past perfect simple were suitable solutions. Students, however, much preferred the past form, and 209 of them provided *found out*. Only 33 students (10%) provided *had found out*, but nine other students tried to form another past perfect form. Nevertheless, it remains somewhat unclear in the context whether the "finding out" took place or not, whether there was anything to find out and, if there was, whether this happened on the day of the murder (see also Section 8.2).

Finally, slot 14 (... **because we ___ (expect) him to come earlier**) seemed to divide opinions. Although many teachers found *had expected* and *had been expecting* more often the best form (but two teachers found them inappropriate), they showed some inclination to accept *expected* and perhaps also *were expecting* (but three teachers found *expected* inappropriate or questionable and five teachers found *were expecting* inappropriate). Thus, there are four potentially suitable forms. Students favoured the past forms: 182 students provided *expected* and 65 *were expecting*, but some students also provided past perfect forms (15%, 47 students). The form *had expected* was given by 44 students and *had been expecting* only by three students. No other past perfect forms were offered, but four students provided a present perfect form, and 16 students provided other past forms.

Indeed, some students' skills in the past perfect would benefit from some improvement, since they did not seem entirely successful at identifying the contexts where the past perfect was

required. It would also be necessary to further investigate why the students were more inclined to provide the form in some of the occasions but not in others.

The past perfect was occasionally offered in slots where it was not expected. It was provided 300 times in 48 slots, which is 2% of all the forms in those slots. In many of these instances, fewer than ten students provided the past perfect, but in a few slots, the past perfect form was more common. Slot 17 (**He and Lucy ___ (obviously/have) a serious row**) had 16 students providing the form *had obviously had*, while two provided *had obviously have* and another two *obviously had had*; three more students provided a past perfect with a variety of spelling errors. In slot 96 (... **but she said she ___ (not/had)**), some students apparently became confused when the word *had* was already provided in the past form in brackets, and 12 students used the past perfect, either with variants of *had had* or *had seen*. In slot 100 (**She ___ (not/even/go) into the study**), 14 students provided six different versions of the past perfect, and in slot 106 (**Lucy ___ (not/walk) to the village and back, if ...**), 12 students produced either *had not walked*, *hadn't walked* or *hadn't walk*. While it is unlikely that in slot 17, the row would be over since Trevor was still shouting, it is somewhat easier to understand that in slot 106, students may have thought that Lucy's walking was over before she met Gerald.

Three slots were even more common with the past perfect than the slots described above: slots 103, 104 and 107. All of these occurred towards the end of the test, and a lot of variation was found in these slots in any case. In slot 104 (**Anyway, Dorothy Stern told her sister she ___ (leave) her husband**), 30 students argued that Dorothy Stern *had already left* her husband and forgot that in slot 31, she told her sister that she had only just decided to do so. Nothing in the story seems to imply that she told her husband about this before he was murdered. However, 26 students provided *had left*, three *had leaved* and one *had levt*. A number of students (40) also provided a past perfect form in slot 103 (**he ___ (leave) the Sterns' house before half past nine**), and some of the teachers had sympathy for this choice. Finally, in slot 107 (**if he ___ (meet) her at twenty to ten)**, where 41 students provided *had met*, *had mett* or *had med*, the choice of the past perfect can perhaps be understood if the students thought that the meeting took place before another event, but this choice is in conflict with the order of events in the story. The fact that the police officers discussed the apparent conflict in times and facts in the story may have confused the students. The last few slots were perhaps difficult to follow if the students had not paid attention to what had been testified earlier in the story.

8.1.4 Present perfect forms

There was only one slot where the present perfect was the only expected form; in some other slots, it could be provided but the present was also accepted. This expected use of the present perfect involved slot 84 (**I ___ (live) here for two years now**), and both the simple and the progressive forms were appropriate in most teachers' opinion. The students were fairly skillful in this, as 254 students provided *have lived*, two *'ve lived* and 27 *have been living*, which already represents almost 89% of the students. Eight more students also produced a misspelled present perfect form. Thus, at least in this one instance, students did not have great difficulty in providing the present perfect in a context where it was expected, while they might also provide it in unexpected contexts.

As we can see in Table 28, students occasionally offered the present perfect in instances where the past perfect would be the standard. However, the extent to which this was manifested differed greatly. The present perfect was provided 227 times in 42 slots where it was not the expected form, which is almost 2% of all forms provided in those slots. The highest number of the present perfect in slots where it was not expected was in slot 31 (**... I ___ (decide) to leave Trevor**), where 26 students offered *have decided* and six other students other present perfect forms. This perhaps reflects a lack of skills in reported speech (see Section 8.1.5). A similar case was in slot 95 (**I asked her if she ___ (see) the dog, but ...**), where 23 students offered a present perfect form, although the past perfect would have been more suitable for reported speech. The form *have had*, provided by 19 students in slot 41 (... **because Dad and I ___ (have) a stupid argument about Alan the day before)**, was also unconventional in the context. The students may have ignored the time reference *the day before* or were perhaps unaware of what to provide with such expressions of time. A similar case was slot 70 (**Lucy ___ (already/leave) the room before ...**): it is not clear why students would interpret this as a case for the present perfect, but 23 students did so. Slot 67 (**Mrs Stern said something about teenage girls and that they ___ problems with Lucy**), however, is more ambiguous, and perhaps the 28 students who provided the present perfect form *have had* thought that the problems began in the past but continued up to the moment in question and beyond; perhaps it could even be interpreted as the normal situation in this family to have problems.

8.1.5 Reported speech

The previous sections have already referred to the fact that the students had some difficulty with tense in reported speech, and this section now reports on the students' success rate in reported speech. In English, there are some changes in word forms when moving from direct speech to reported speech, and tense is one of the elements that is likely to change. According to *English Grammar in Use* (Murphy 2004, 94), "the main verb of the sentence is usually past" in reported speech, and the tense in the reported content typically changes. You may either keep the past tense from direct speech the same in reported speech or change to the past perfect, but if the situation in the reported content has not changed in the meantime, you may keep the same form as in the direct speech (Murphy 2004, 94-97). In Finnish, reported speech follows different conventions and often leaves tense unaffected or ambiguous (see Section 5.1.1).

There were ten instances in the test where verb forms in reported speech were required: slots 7, 8, 31, 54, 75, 76, 77, 95, 96 and 104, and they were indicated in the test with, for example, "*After dinner last night he said he ...*" (slot 7). The reporting verbs *tell*, *ask* and *say* were used in these contexts. It is somewhat unclear whether slot 8 (**He ___ (have) a meeting with Gerald, his business partner, the next morning**) is reported speech, with the same reporting verb as in slot 7 (**After dinner last night he said he ___ (want) to check some business papers in his study)**, or whether only slot 7 is the reported speech, after which there is an explanation about why Trevor Stern needed to check his papers. The reporting verbs that were given in the test were all in the past tense, but in two slots, students were asked to provide the correct form of the reporting verb as well; this was in slots 30 and 53. The context around slot 30 seemed to expect the past tense and that around slot 53 the past perfect, but some students did not provide these forms and some teachers also accepted the past tense in slot 53.

The forms that the students used in these slots are provided in Table 29. It only contains the instances where the expected form was used and the corresponding forms in other tenses (for example, the corresponding present form of *wanted* is *wants*; the corresponding present perfect for *had decided* is *have decided*). Only words that were spelled in the standard way are included. Slots 8, 54 and 104 are not given in the table because teachers disagreed on the expected form; these are discussed separately below and in Section 8.2.

Table 29. Forms used in slots with reported speech when compared against the expected forms, with the figure for the expected form given in bold

Slot	Standard expected form	Expected or corresponding past form	Expected or corresponding past perfect form	Corresponding present form	Corresponding present perfect form
7	wanted	**265 (83%)**	-	31	-
31	had decided	78	**186 (58%)**	4	27
75	was	**229 (72%)**	5	61	2
76	did not know / didn't know	**231 (72%)**	-	28	-
77	was talking	**196 (61%)**	-	23	-
95	had seen	38	**218 (68%)**	1	8
96	had not / hadn't	**225 (71%)**	8	6	-

The students were fairly comfortable with using the reported form in the past tense, although some students used the present forms as well. In slots 7, 75, 76 and 77, many students offered the present form, and in slots 31 and 95, many students offered the past form instead of the past perfect. Actually, slot 75 (**He said I ___ (be) an ignorant country doctor)** could be interpreted to convey a timeless fact instead of a single event, and therefore the use of the present tense in reported speech can perhaps be justified. Most students seemed aware of the implications of reported speech on tense, but a third of them would have benefitted from further practice. The rate of success in these forms was from 58% to 83%, with an average of 69% (221 students).

Although I refer to slots 8, 54 and 104 in Section 8.2 in relation to the interpretation of time, they were also cases where reported speech was used, now in the context of referring to a past event where, from the perspective of that point of time, a future event was discussed. Thus, the reported speech form could contain a so-called past future form (also called "the future seen from the past", e.g. in Hewings 2005, 28). The verb forms used could include, for example, *was going to have*, *was having*, *would have*, *would be having*, *was to have* and *was supposed to have*. The past future forms used in these slots are given in Table 30. Only forms with standard spelling are included in the table. The table also provides the figures for using the past simple.

Table 30. Past future and the past simple in students' responses

Slot	would / 'd	would be -ing	would have / would've	was going to	was -ing	was to	was supposed to	was about to / was planning to	should	Past simple
8	28	0	22	15	3	4	1	1	1	219
54	61	21	0	10	107	1	0	0	1	47
104	70	0	0	27	70	1	0	1	0	44

Interestingly, the students were noticeably more likely to treat slots 54 (**... he ___ (play) in a concert that evening**) and 104 (**Anyway, Dorothy Stern told her sister she ___ (leave) her husband**) as instances including a past with a future reference, while the past simple was preferred in slot 8 (**He ___ (have) a meeting with Gerald, his business partner, the next morning**). This may be because in slots 54 and 104, reported speech is more clearly indicated in the preceding sentence. However, slot 8 was the only slot where the future time was clearly indicated in an expression with a specific point of time in the future, "*the next morning*". It is somewhat surprising that students did not more readily interpret the sentence as referring to a future intention, which actually even remained an intention, because Trevor Stern died before the intended meeting. For some reason, students did not seem to pay attention to the contradiction in the sentence if they used *had* in slot 8.

In slot 8, past future forms were provided by 75 students (24%). In slots 54 and 104, a more convincing number of students created a link to a future time with a past form, with 201 students (63%) using a past future form in slot 54 and 169 students (53%) in slot 104. In these three slots, a number of students again used the present: these students may have been ignorant of the implications of reported speech, or they may not have been able to apply the knowledge they had. Furthermore, instruction and practice could perhaps be recommended.

8.2 Interpreting event time

As referred to above, a number of slots in the test were particularly likely to be interpreted in a variety of ways as far as the choice of tense and aspect is concerned. This resulted in significant variation in the student responses, and the teachers reacted to such variation in different ways. It seems that in a number of these instances, some teachers acknowledged the fact that the event time was ambiguous, but some other teachers seemed not to have realised that such slots could be understood in a number of ways, or they may have found the slots unambiguous.

For example, slot 8 (**He ___ (have) a meeting with Gerald, his business partner, the next morning**), which is actually an instance of reported speech referring to a future event that we now know did not happen, invoked a number of interpretations. While the majority of students (219) provided the simple past *had*, there were also students offering the past progressive and a variety of forms referring to the future, as well as forms in the conditional. Thus, some students simply seemed to go for past marking, without perhaps even realising this is reported

speech, while others had focused on the phrase '*the next morning'* at the end of the sentence after the verb, and perhaps tried to offer a future reference. It remains unclear, of course, how carefully students read the entire sentence and how conscious they were of the implications of the fact that this was a form with reported speech (see Section 8.1.5). Nonetheless, some students showed awareness of the fact that since Trevor Stern died, he could not go to the meeting he was supposed to by choosing, for example, *was to have* (4 students), *was supposed to have* (1 student) or *would have had* (21 students). Teachers were to some extent undecided here, but most support was given to *was going to have*, *was supposed to have* and *was to have*, and several also accepted *had*.

Slot 54 (**... he ___ (play) in a concert that evening**) was also an instance of reported speech, and students were undecided on what tense would best suit the context. While some simply selected the past tense (107 students with *was playing* and 47 students with *played*), a number of students seem to have sought a form that would help match the fact that the telling that was referred to had happened previously, and the boyfriend had thus probably used a future reference in his direct speech. While we do not know what he had actually said, students offered, for example, *would be playing* (21 students), *was going to play* (10 students) and *would play* (60 students). Altogether, 34 different forms were given. The teachers accepted a number of these forms, but only two teachers accepted *played*.

A variety of forms were also given in response to slots 39 and 40 (**I ___ (never/like) Mum or Dad to be around when I ___ (talk) to him**), which are also discussed in Section 8.5 from another perspective. Students gave a range of interpretations to the situation in slot 39, depending on whether they thought the "not liking" in question should be marked with the present, past or present perfect. On the one hand, it referred to a (recurring) instance from the past, since although the father was now dead, the mother was still alive. On the other hand, it could be interpreted as a timeless statement, something that would hold at any time; this would require the present tense form. A further interpretation is that this was something that began in the past but continued to the present, which would call for the present perfect. While the most popular form given by the students was *never like* (107 students), it was closely followed by *never liked* (101 students), and the form *have never liked* was also supported (82 students). The teachers greatly disagreed on this, and each of the forms quoted above was accepted by several teachers as the best or acceptable and dismissed by some others as inappropriate. This truly seemed to divide the teachers. The book offers only one solution, *never like*. Slot 40 is part of the same sentence as slot 39, and it is not surprising that the variety in forms continued here.

Students were torn between the present and the past tense, but also between the simple and the progressive. The present tense was more popular, with 131 students providing *talk* and 63 *am talking*, but there was also support for the past, with 59 students for *talked* and 50 for *was talking*. Teachers also displayed variation in their assessment, again with some teachers supporting each of the forms mentioned and rejecting some other forms.

Another slot evoking different interpretations about the time of the event was slot 67 (**Mrs Stern said something about teenage girls and that they ___ (have) problems with Lucy**), where all tenses received some support. While the most common form was *had* (149 students), a number of students also supported the forms *have* (35 students), *have had* (28 students) and *had had* (39 students); furthermore, some students considered this a progressive event, offering *were having* (42 students) and *are having* (7 students). The teachers clearly preferred three of these forms: *were having*, *had* and *had had*, but there was some disagreement as well.

Slots 83 (**Yes, my house ___ (be) just round the corner from the Sterns'**) and 85 (**I ___ (have) a little cottage in the village)** also involved decision-making: how many dwellings does Gerald have? In slot 83, the majority of students believed that his present house *is* round the corner from the Sterns' (239 students), and some believe it *was* there (71 students), but the main disagreement concerned the cottage in the village in slot 85: does he still have it? The result is almost a tie: 152 students argued he *had* the cottage, while 151 students thought the form to use was *have*. Teachers had more faith in the form *had*, but a number of them also accepted *have*.

Another decision regarding event time had to be made about the potential "finding out" in slot 101 (**And perhaps Trevor Stern ___ (find out)**): if it happened, was that on the day of the murder or before that? Altogether, 209 students thought the form *found out* was needed here, but 33 students thought *had found out* was better. Most teachers accepted both forms.

Students also understood slot 104 (**Anyway, Dorothy Stern told her sister she ___ (leave) her husband**) in a variety of ways. Some believed the wife had already left her husband (*had left*, 26 students) and some argued she did it that particular day (*left*, 44 students), but most students thought this was something that would have happened in the future. Twenty-nine different forms were produced, including *was leaving* (70 students), *was going to leave* (27 students) and *would leave* (69 students). The teachers were more inclined to think that this was still in the planning stage and gave their support to the forms *was going to leave* and *was leaving*; furthermore, many but not all the teachers accepted the conditional form.

There was also time disagreement with slot 99 (**It ___ (not/be) his wife)**, where some students supported *was not* / *wasn't* (56 and 127 students), and others *is not* / *isn't* (36 and 28 students). Other students used a modal auxiliary and created a variety of forms involving *can* and *could*. The teachers were torn between these forms, and there was some understanding for at least 12 different forms.

The majority of the slots presented above readily lend themselves to various interpretations, and it is therefore difficult to argue for one single correct answer in these instances. Thus, whether a particular form was accepted or not depended ultimately on the relevant teacher's understanding of the context and willingness to seek and accept alternative ways to rate the verb forms offered for the slot.

8.3 Modal verbs, irregular verbs, passive forms and verb complementation

This section explains the extent of variation that can be witnessed in the students' use of modal verbs (Section 8.3.1), irregular verbs (Section 8.3.2), the passive (Section 8.3.3) and verb complementation (Section 8.3.4). These are phenomena that arose from the results in an interesting way.

8.3.1 Modal verbs

The test provided no instruction on whether students were allowed or expected to use modal verbs in their responses. In this study, very few students did so early in the test, but many felt drawn to modal verbs in some of the slots towards the end of the test. No student used a modal auxiliary in more than six slots, and some never used them. The book key gives the correct answer with a modal auxiliary in five slots: slots 49, 74, 99, 103 and 106. In slot 49 (**He ___ (not/see) me, though, because it was dark outside**), only one student used the form *couldn't see*, which the book key acknowledges as the expected answer together with *didn't see*. The use of the word *would* in slot 74 is discussed in Section 8.7. Both *can* and *could* are possible in slot 99, as discussed in Section 8.2.

Students mainly used modal auxiliaries in slots 103 and 106, although the numbers remained small. In slot 103 (**... he ___ (leave) the Sterns' house before half past nine**), four students used a modal auxiliary and produced *must have left*, *would have left* and *should have left*. Of the forms, teachers particularly liked *must have left*; they were more cautious with the other two modal forms. In slot 106 (**Lucy ___ (not/walk) to the village and back, if ...**), both *can* and *could* were used in altogether seven different forms. Teachers approved of their use particularly in *could not have walked* / *couldn't have walked*, and some also in *cannot have walked* / *can't have walked*. However, only 20 students used verb forms with a modal auxiliary in this slot.

Students did not use modals very extensively elsewhere, either. In slots where the book did not expect a modal auxiliary, *may* and *shall* did not appear at all, while *might* and *must* were used in one slot each. *Should* appeared in six slots, *can* in 3 slots and *could* in 5 slots. Only *will* and *would* were used more extensively, *will* in 17 slots and *would* in 22 slots. This may be because students did not come to think of the possibility of using modals in the test for other purposes than referring to the future or expressing a condition. The use of modals is visualised in Table 31 below.

Table 31. The use of modal verbs in the study

Modal verb	Number of students	Number of verb forms	Number of slots
may	0	0	0
might	1	1	1
must	2	1	1
shall	0	0	0
should	7	6	6
can	28	10	3
could	44	14	5
will	101	18	13
would	293	39	22

8.3.2 Irregular verbs

Students mainly seemed to know how to use irregular verbs and only a few misspelled them (see also Section 8.2). In total, a regularised form of an irregular verb was used in 415 instances in 23 slots, which is about 6% of all the forms provided in those slots. For example, *told* was provided correctly by 312 out of the 319 students in slot 11 and *took* by 299 students in slot 18. Two students provided *telled* in slot 11 (and by slot 30, one of them had already learned to use

told; a few students produced a form that looked like an attempt to produce *told* but they misspelled the word) and five *taked* in slot 18 (in slot 43, three new students used *taked* and one of these students had learned *took*). One student spelled *comed* in slot 24, while 278 used *came*. The past form *let* was created by 300 students in slot 63, and 302 students produced *bought* in slot 86, with 6 students writing *buyed* and one *buyd*. A vast majority of students, 307, used *met* in slot 92, with only one *meeted*. *Found* was given by 300 students in slot 98 as opposed to three students with *founded* and five with *finded*. Three students gave *meaned* in slot 102.

According to Finegan (2004, 548), children learning English as their first language most frequently overgeneralise the regular ending to the following verbs: *eated*, *maked*, *finded*, *hitted*, *falled*, *doed*, *speaked*, *breaked*, *goed* and *runned*. Of these, only *finded* was frequent in this study; *goed* or *doed* were not provided at all, and there were no instances of *get(t)ed* or *maked* either. The rest of the words on Finegan's list (2004, 548) were not required in the test. Some of the verbs that are irregular in standard English may be regular in some dialects or regional varieties of English; for example, *seed* is used in some southern American dialects instead of *saw*, while *seen* is used to replace *saw* in the north (Leith 2007, 137), and *went* is an appropriate past participle of *go* in Irish English (Rhys 2007, 209). Rhys (2007, 211) further maintains that some native speakers are also unsure about the distinction between the past tense and the past participle in some irregular verbs. However, whether the Finnish students who provided such forms were aware of these dialectal forms cannot be confirmed.

A few problems with irregular verbs did occur, for example with the verb *see*. The form *saw* was produced by 303 students in slot 45, but it was slightly misused in *did saw* as well. One student, however, formed *sawed* in slot 47 and seven students wrote *had sawn* and 17 *had saw* in slot 95; there was also one *was saw* and one *has been saw*. This indicates that some students were confused with the irregular verb *see* and the regular verb *saw* or failed to distinguish between the past simple and the perfect participle form. Regarding the verb *hear*, 27 students formed the past tense as *heared* in slot 23 as opposed to 283 students with *heard*. Of these 27 students, 22 made the same error in slot 26 together with four new students, who originally produced the correct form. In addition to *stood* in slot 48, 26 students spelled *standed*, and in slot 70, there were 33 students attempting to use *leaved* either alone or in various multi-word combinations, such as *had already leaved*; in slot 80, *leaved* was only given by 23 students. Furthermore, although 273 students created *lost* in slot 90, there were also 26 students with *losed*, six with *loose* and four with *loosed*. A number of students (10) also struggled with the

word be, using *didn't be*, *did not be* and *did'nt be* in slot 82 instead of *weren't* or *were not*. The case of *spill* is discussed below in Section 9.2.2; *creep* is discussed in Section 8.7.

At times, the uncertainty some students had with irregular forms was reflected in their marking of the past time twice, either by adding *–ed* after the irregular form (as in *tooked*) or by using the irregular form after *did not* or *didn't*, or sometimes the emphatic *did* (as in *didn't murdered* or *did saw*). This double past marking occurred 147 times in 22 slots (2% of all the forms in those slots). The most common was *didn't saw* in slot 49 (where the expected form is *didn't see*), which was provided by 22 students; *did not saw* was given by two students and *didn't seen* by one student. In slot 96, 13 students provided *didn't had* or *did not had*, and 14 students provided either *did not even went*, *didn't even went*, *didn't even gone* or yet some other form with two markings for past tense in slot 100. The confusion with the forms of *be* was also evident in slot 82, with *didn't were* or *didn't was* given by six students. Double past markings occasionally also affected regular verbs. Examples are *didn't murdered* and *did not murdered* in slot 6 (20 students in total) and *didn't answered*, *did not answered*, *didn't aswered* and *didn't answerd* in slot 51 (13 students in total); even *doesn't answered* was suggested.

Thus, although the majority of students mastered the use of irregular verbs, a handful seemed uncertain about them. It cannot be concluded why the verb *see* caused such difficulty, unless this reflects some confusion with the regular verb *saw*. In addition, a small number of students struggled with the verb *be*, which is somewhat surprising given its frequency. However, some forms of the verb *be* bear little similarity to the infinitive, and the fact that *be* is also used in a variety of ways as an auxiliary may have confused the students.

8.3.3 The passive

There were only two slots in the test where the passive was supposed to be used, although some students attempted to create a passive elsewhere as well. However, the students had great problems forming the suitable passive form in slot 108, although they scored better in slot 4. This is perhaps because the required form in slot 108 (**She ___ (still/shout at) by her father at nine-fifteen)** was the past progressive form combined with the adverb *still* and preposition *at*, which turned out to be very challenging for the students. Only 28 students managed to supply the expected form, *was still being shouted at*. Forty-eight students managed to produce a suitable passive form but without the progressive, and 102 formed the progressive without the

passive, while another 48 did so as well but ignored the preposition. The students suggested a total of 44 different forms for this slot, which was the greatest number of variant forms produced per slot (tied with slot 100).

Students succeeded better in slot 4 (**It was soon clear that he ___ (murder)**), where 68 students created *had been murdered* and 173 provided *was murdered*. Given that 44 students provided no answer to this slot, the percentage of students who successfully produced a passive form was much greater in this slot, particularly since 18 more students produced a form that looked like an attempt to create a passive form. In slot 4, most students managed to provide some attempt at a passive, with only about 10 students providing an active sentence (some students provided forms that are difficult to categorise). In contrast, about 210 students provided an active-looking form in slot 108. It seems, indeed, that students found slot 4 much easier to identify as a slot requiring the passive than slot 108. This may be because of the other aspects of verb usage that students needed to remember and apply in slot 108. Thus, perhaps the combined challenge of using the progressive with the words *still* and *at* made slot 108 more difficult for the students.

The Swedish-speaking and bilingual students were slightly better at identifying the need for a passive in these two instances, but the difference was very small. In slot 4, most of them (19 out of 21) provided a passive either in the past perfect or the past, while in slot 108, the bilingual and Swedish-speaking students provided ten different forms, and only three students (out of 21) provided the expected form, *was still being shouted at*. However, six Swedish-speaking and bilingual students were able to form a passive without the progressive, which is proportionally more than Finnish students. It seems that the Swedish-speakers were not fully able to benefit from the similarity in the structure of the passive in Swedish and English. However, this may be because Swedish has several structures to choose from to provide a passive, and only one of them resembles that used in English passive formation (see Section 5.1.2).

The students in this study were not likely to use passive forms when the active was expected, but there were a few instances where the passive was used in these situations. This happened 67 times in 27 slots, a mere 1% of all the forms provided. Most of these were single instances or given by two to four students, but eight students provided a passive form in slot 31 (*was decided*, *has been decided*, *had been decided*) and six students in slot 61 (*was planned*, *was planed*), while as many as 18 provided a passive in slot 59 (*was called*). Although it is somewhat difficult to understand why another person would decide for Dorothy in slot 31 (**I ___ (decide) to leave Trevor**) or why another person would plan on behalf of Doctor Emerson in slot 61 (**...**

rather later than I ___ (plan) to be because ...), the passive in slot 59 (**I ___ (call) at the Sterns' house at nine-fifteen**) can be understood if students were not familiar with the word *call* (*at*) in the sense of '*visit'* but only in the sense of *'phone'*, in which case the passive makes a little more sense.

8.3.4 Verb complementation

There were two sources of variation in verb complementation in this test, with how to continue a sentence with a second verb after a particular main verb has been given. Typically, this means choosing between the infinitive with or without *to*, using the gerund (the *–ing* form) or using a subordinate clause, for example with *that*. English verbs tend to follow a particular pattern, with some verbs only allowing one of the above ways of complementation, while other verbs are more versatile, but occasionally have a difference in meaning between two types of verb complementation. This phenomenon is also discussed in Section 9.2.2, with comments on a change in contemporary usage that has increased verb complementation with the gerund in this type of structures.

According to *English Grammar in Use*, in structures such as *make somebody do something*, you do not use the infinitive marker *to* before the infinitive (Murphy 2004, 110). The majority of the students (172) seemed to be familiar with the rule and produced *take* in slot 33 (**... so she made me ___ (take) Dad's tea into his study)**, but 100 students provided *to take*. Only one of the teachers was willing to accept *to take* and one found such use questionable; the others found the form inappropriate.

The verbs *try* and *start* can be followed with either the infinitive or *–ing*, with a slight difference in meaning. After *try* in slot 72 (**I tried ___ (explain) to Trevor why ...)**, students offered *to explain* much more often (230 students) than *explaining* (34), and some students (45) also suggested *explain* without *to*. Teachers were unanimous with accepting *to explain* as the best form, but only 6 treated *explaining* in the same way; others either found it acceptable (2 teachers), questionable (one teacher) or inappropriate (4 teachers). After *start* in slot 87 (**... when I started ___ (earn) a lot of money**), two-thirds of the students (200) provided *to earn* and one-third of the students (96) *earning*. Teachers found *to earn* the best form and considered *earning* either the best or acceptable (apart from one teacher). I return to this issue in Section 9.2.2 to discuss the patterns of change observed in the use of *start* by native speakers.

8.4 Unusual forms

Another aspect to analyse is whether a particular student or student group was mainly responsible for unusual forms, which here are defined as forms with 30 students or fewer giving a specific response, accounting for less than 10% of the entire student population. Some of these forms were considered acceptable (or the best) forms, i.e. they are grammatically appropriate forms, which are just rare in use (at least in the particular context), but the majority are inappropriate forms, including spelling errors, unconventional tense formations and uninflected forms; in addition, the category "unusual forms" includes cases of no answer[88]. There were a total of 3973 such unusual forms, which corresponds to a mean of 12.45 per student (SD = 10.04). The results are presented in Table 32. The distribution of the unusual forms, however, is uneven so that the mean for English Majors is 5.83, while the mean for all the other groups is over 10, with the highest mean of 22.08 for the Visual Arts students. Furthermore, the range in the number of unusual forms is different for the different groups, with the English Major group again being very different from the rest of the groups. The highest number of unusual forms for English Majors and minors is 11, and there are two students with no such forms, while the other groups have as many as 30 (Visual Arts), 46 (both Computer Science and Technology), 50 (Humanities) and 66 unusual forms (Grammar group).

Table 32. The distribution of unusual forms (used by 30 students or fewer) in the data

Student group	Total number of unusual forms	Mean per student	Standard deviation	Lowest number	Highest number
English Majors (n=35)	204	5.83	3.12	0	11
Grammar (n= 42)	736	17.52	14.21	1	66
Humanities (n=31)	354	11.42	9.01	1	50
Visual Arts (n=12)	265	22.08	5.74	9	30
Computer Science (n= 17)	302	17.76	15.27	2	46
Technology (n= 182)	2112	11.60	8.32	1	46
total (N=319)	3973	12.45	10.04	0	66

[88] This category includes 58 slots of no answer, and since only one of the 'no answer' slots exceeds 30, it is included in the analysis. This is slot 4, with 44 missing answers.

Thus, it seems that among unusual forms, the students with the strongest command of standard acceptable forms were students of English Philology, followed by students of Humanities and Technology. The greatest number of unusual forms weas given by students of Visual Arts, Computer Science and Grammar group students: they created a multitude of innovative forms that are not considered standard. It might be, of course, that there were more students with dyslexia in these groups, or these students might not have been as motivated to learn the standard spelling as some other students. Unfortunately, this information is not available for analysis[89].

An analysis of the unusual forms used by the English Majors shows that most of the forms were simply less common tense choices. Spelling errors were rare, but some forms were non-standard, such as *crep* for slot 37, *heared* for 23 and *have went* and *had went* for 20. Some problems arose from ignoring spelling changes when inflecting the word, as in *stoped* in 68 and *argueing* in 79, or misinterpreting or overgeneralising such rules, as in *was lookking* in 58 and *was still beeing shouted at* in 108. However, the vast majority of the forms were well-formed; they were either just rare but acceptable forms in the context or grammatical but inappropriate forms in the context. Students in the other groups had spelling errors, non-standard forms, misinflected forms and also forms that were well-formed per se, but not always appropriate in the context. In these cases, the fact that a student had a vast number of unusual forms typically means that the student made many errors.

Spearman's rank order correlation between the number of errors (the lenient approach) and the number of rare forms is surprisingly high (r_S = .873, p < .001). This indicates that although some of the rare forms are actually acceptable forms, the vast majority are not. The provision of rare forms can be student-specific: if a student had a tendency to misspell words, this individual's score could quickly increase the group mean. A student who had internalised a non-standard way of forming a particular tense could do the same. Therefore, the next section investigates, among other issues, whether this is a common phenomenon: are students systematic in providing unusual forms?

[89] Four students (Student 42, Grammar; Student 94, Humanities; Student 181, Technology; Student 201, Technology) self-reported on dyslexia. They had, respectively, 16, 50, 12 and 11 instances of unusual forms. Thus, only one of the dyslexic students (Student 94) stands out; the others do not differ from the rest of the population. Of course, there might have been other dyslexic students who did not report on this.

8.5 Intra-student consistency

This section explores how systematic certain phenomena were in student responses. Regan (2013, 277; see Chapter 2) argues that L2 variation is systematic, although it can contain "some random developmental variability in learner language". In this study, it seems that some students were consistent in making the same types of errors in different slots and these can, then, be considered to be fossilised (see Section 3.2.1). Other students made occasional errors but provided more standard forms at other times, which seems to indicate that the errors were just slips or potentially results of misreading their handwriting[90], for example. Unfortunately, the distinction between an error and a mistake or between local and global errors (see Section 3.2.1) cannot be made on the basis of the responses here.

Some students were, indeed, systematic in producing similar types of forms in a number of contexts, although others only seemed to notice the existence of such contexts in the most prototypical cases. Some students were careful in following consistency in time (i.e. tenses), while others seemed less concerned with systematic behaviour in this aspect of language. In this section, I focus on three such phenomena: 1) the provision of unusual spelling and unconventional forms of some irregular verbs, 2) the use of contractions or lack thereof, and 3) tense consistency within a sentence. However, it is sometimes difficult to conclude whether unusual spelling is a spelling error or a non-standard formation of a verb form.

First, there was some systematic provision of unusual spelling and unusual irregular forms. Student 10 was systematic[91] in doubling the final *t* in the word *shout*, thus producing *was still shoutting* for slot 16, *shoutting* for 19, *did not shoutt* for slot 27, *shoutting* for slot 66 and *was still shoutting at* for slot 108, while Students 35, 141 and 170 did so only once. Similarly, Student 47 used the form *heart* for slots 23 and 26, while two students who first produced *heared* for slot 23 changed to *heard* by slot 26, and three students who first produced *heard* for slot 23 changed to *heared* by slot 26. However, 22 students gave *heared* to both slots. Furthermore, Student 107 began by spelling *did'nt* but changed to *didn't* by the fourth instance of this form. Student 53 used *told* in slot 11 but *telled* in 30, while students 57 and 117 did the opposite, and they all used *told* in slot 53. Student 153 was systematic in using *gick* as the past

[90] For a discussion of this, see Section 11.2.

[91] Interestingly, this student did not double the *p* in *planed* in item 61. This seems to indicate that the student was confused about the rules for doubling consonants before word-final morphemes.

tense of *go*, which can be explained with crosslinguistic influence from Swedish: the past tense of the Swedish word for *go* (*gå)* is *gick*. This student was the only one to use the form *gick*. Student 125 used the correct form *saw* as the past tense of *see* in slot 45 but slipped to *sawed* in slot 47, while Student 67 was systematic with *sow*. However, Student 67 later produced *didn't see*, while Student 125 was consistent with *didn't saw*. Furthermore, Student 67 used *meat* in slots 56, 92 and 107 as the past tense of *meet*, while Student 110 systematically provided the forms *mett* and *had mett*. This form may also be influenced by Swedish: the past tense of the Swedish word for *meet* (*möta*) is *mötte* and the past participle is *mött*.

The second topic to discuss for consistency is the use of contracted forms. Interestingly, the students in this study rarely provided contracted forms in the affirmative, i.e. verb contraction (see Biber et al. 1999, 1128). Table 33 shows the contracted and uncontracted forms pairwise, and it only includes forms where at least one student used a contracted form and where the sentence was intended to be affirmative. As we can see, contracted forms are only used in the affirmative in the forms *'d, 'm*, *'ll* and *'ve*, but their total number is 18, while their uncontracted counterparts in the same contexts are given 586 times; thus, the contracted forms occur in a mere 3% of the cases. No student used contractions systematically. Only one student (Student 90) used contracted affirmative forms four times, while Student 47 did so three times (each with *'ll*) and Student 135 twice (also with *'ll*); the rest are single usages of contracted forms in the affirmative. Thus, at least in this study, students were not likely to use contractions in the affirmative.

Table 33. Contracted and uncontracted forms in the affirmative, given pairwise

Slot (forms)	Uncontracted[92]	Contracted	Total
8 (would have had / would've had)	21	1	22
31 (have decided / 've decided)	26	1	27
40 (am talking / 'm talking)	63	3	66
54 (would play / 'd play)	60	1	61
54 (will play / 'll play)	13	1	14
60 (would be / 'd be)	18	1	19
60 (will be / 'll be)	1	2	3
84 (have lived / 've lived)	254	2	256
104 (would leave / 'd leave)	69	1	70
104 (will leave / 'll leave)	29	1	30
110 (will make / 'll make)	32	4	36

[92] Double forms are excluded because in each of these cases (4), one form is contracted while the other is not.

The situation changes when we explore the negative forms. Table 34 shows the distribution of the use of contracted and uncontracted forms in negative sentences (i.e. the *not* contraction; see Biber et al. 1999, 1128). When the expected response was the negative past simple, i.e. a form beginning with *did not* or *didn't*, students frequently used the contracted form. In such slots, over 70% of the students used the contracted form. Students were fairly systematic in their choice, but a small change occurred between the two instances of the expected form *didn't want* vs. *did not want*: 13 students more used the contracted form in slot 36 than in slot 22. In the other slots, however, the difference was marginal. When the form to be contracted was not a form of the word *do* but of *be* or *have*, the percentage of using contracted forms decreased a little. In slot 27, where the expected form was the negative past progressive, 70% of the students still used a contracted form. With *hadn't* in slot 96, the score was 69%, but it dropped to 61% for *wasn't* in slot 99 and 57% for *weren't* in slot 82.

Table 34. Contracted and uncontracted forms in the negative

Slot (contracted form)	**Any contracted negative forms**[93]	**Any uncontracted negative forms**[94]	**Affirmative, not contractable, undecided or no answer**
5 (didn't love)	231	86	2
6 (didn't murder)	232	86	1
22 (didn't want)	236	80	3
36 (didn't want)	252	67	0
49 (didn't see)	244	74	1
51 (didn't answer)	245	73	1
74 (didn't let)	242	75	2
76 (didn't know)	249	69	1
100 (didn't even go)	239	69	11
106 (didn't walk)	241	71	7
27 (wasn't shouting)	225	90	4
96 (hadn't)	219	96	4
99 (wasn't)	195	116	8
82 (weren't)	183	130	6

This implies that students were well versed in using the negative contracted form *didn't*, but were either a little less confident with contractions using *have* and *be* or had perhaps received less exposure to these words in the contracted form. Since contractions were rare with the

[93] Thus, in addition to the "default" contracted negative form, any other form that appears in the contracted negative form is included here. This could be *didn't*, *can't*, *isn't*, *won't*, *hadn't* etc. Whether the form is correctly spelled or grammatical plays no role here, as long as it can be identified as an intention to create a contracted form.

[94] In addition to the "default" uncontracted form, any form that appeared in an uncontracted form (with *not* spelled as a full word) was included in cases where the contraction would have been possible.

affirmative, where such a form only appeared with forms of *be*, *have*, *would* and *will* / *shall*, the explanation may indeed be that students had more frequently encountered models of using *didn't* but met other words in the contracted form less often. Another possibility is that students had been discouraged from spelling words in the contracted forms in academic contexts, but this would not explain why negative forms were so frequent in the contracted form. Corpus findings with native speakers indicate that contractions are common in spoken language and in written genres reporting on spoken discourse, such as conversation and fiction, but less so in news and academic texts (Biber et al. 1999, 1128-1132). The frequency of contractions is typically dependent on the next verb and on the nature of the subject before them: contractions in the affirmative are more likely with pronoun subjects and when the contracted verb is an auxiliary (Biber et al. 1999, 1128-1132).

Students were quite consistent in their choice of either using or not using negative contractions. Out of the 319 students, 34 never provided a contracted form and 14 only did so once, while 55 did so in all but one case and 105 students did so systematically. This is described in Table 35. Interestingly, only some English Majors, Grammar students and Technology students had no contracted forms at all, while the Humanities, Visual Arts and Computer Science students can mainly be found at the end of the scale where contracted forms were frequently used. The students who used contracted affirmative forms also used negative forms frequently, in 10 to 14 instances.

Table 35. The number of students using a contracted negative form

Contracted negative forms	Number of students
0	34
1	14
2	7
3	4
4	4

Contracted negative forms	Number of students
5	4
6	3
7	4
8	7
9	7

Contracted negative forms	Number of students
10	10
11	26
12	35
13	55
14	105

The third element regarding consistency to discuss here is the systematic marking of time. As it would be very impractical to track all the slots, I focus on two particularly promising sets of slots to see if students were systematic in their choice of tense: slots 39 and 40 and slots 75, 76 and 77. These are chosen because of the extent of variation found in them but also because of the fact that they are part of the same sentence and they should not therefore really show extreme variation within the same student. The variation in the slots within the same sentence should

not be random within an individual; rather, some continuity and consistency in tense choice would be expected.

Slots 39 and 40 belong to the same sentence (**I ___ (like) Mum or Dad to be around when I ___ (talk) to him**) so that slot 39 is in the main clause, which is followed by a subordinate clause containing slot 40, and they are connected with the conjunction *when*. A variety of forms were given to both slots, with suggestions including present, past, present perfect and past perfect forms. Of the 203 students who provided either the present or present perfect tense in the first slot, 180 students provided the present tense in the latter slot, while two provided the present perfect and six students provided a form where the intended tense cannot be verified[95]. Fifteen students provided a tense mismatch and used the past or past perfect in the second slot. Thus, 89% of the students were systematic with the present forms. Of the 115 students who first provided either the past or the past perfect, 96 also provided a past form in the second slot; no students provided the past perfect. Nineteen students provided a tense mismatch and used the present or present perfect in the second slot. This means that 83% of the students kept using past forms throughout the sentence. The percentage is slightly lower than that of present tense usage.

Slots 75, 76 and 77 are also part of the same sentence (**He said I ___ (be) an ignorant country doctor who ___ (not/know) what he ___ (talk) about**), and they are all parts of a subordinate clause which is indicated as being the contents of reported speech. The main clause is in the past tense, and thus it would be logical for students to report the full contents so that they persist with the same tense they used for the first slot[96]. The analysis demonstrated that 205 students used the past tense and 30 students the present tense in all the three slots. In 46 cases, students mixed the past and the present tense, and 17 students used another combination of tenses, such as the past perfect, the future with *will*, a conditional with *would* or an infinitive with or without *to*; in 21 cases, the intended tense cannot be verified (if the student only provided e.g. *talking*, if there was no answer in the first slot or if a double answer was given). Thus, the consistency rate in these three slots is 74%.

[95] In addition, one student did not provide an answer to slot 39.

[96] However, it can also be argued that the first verb might be a more permanent state of affairs and the second and third context-bound, which would mean that the tense in the first verb might differ from the second and the third.

8.6 Other variation phenomena

This section presents a selection of other phenomena that arose from the data in an interesting way and are therefore worthy of discussion[97]. These include 1) fixed phrases, 2) word order, 3) dropping words, 4) changing the verb to be used, 5) adding words and 6) the emphatic *do*.

Some of the slots required knowledge of the way a particular set expression or phrase is typically used, and it is evident that without knowing how the particular structure behaves, a deviant answer can easily be supplied. Two such expressions were given in the text, in slots 79 and 110. Slot 79 (**... it was no use ___ (argue) with him so ...**) was preceded by the expression '*it was no use'*, and 101 students seemed aware that they were expected to continue this with the *–ing* form and produced *arguing*. The majority, however, seemed unaware of this, and 158 students provided *to argue*. In slot 110 (**I think it's time we ___ (make) an arrest**), the phrase used just before the slot was '*it's time we'*, and only 49 students provided the expected form after such an expression, *made*. Altogether, 200 students provided *make*, 13 students provided *to make* and a number of students provided a form referring to the future, such as *will make* or *are going to make*. Thus, it seems that the majority of students were not familiar with either of these expressions or were not able to apply the rules they may have learned. These two expressions are further discussed from the teachers' perspective in Section 9.2.2.

Some students seemed confused with word order when they were expected to include an adverb in the verb phrase. Such adverbs were to be used in eight slots and included *still* (slots 16, 25 and 108), *obviously* (slot 17), *never* (slot 39), *normally* (slot 42), *already* (slot 70) and *even* (slot 100). The majority of the students (189) in slot 16 correctly placed *still* in the middle of the progressive verb form with *was still shouting*, but both *still was shouting* and *was shouting still* occurred, albeit only given by three students in total. Thirteen students also placed the word *still* after a single main verb in *shouted still* and four in *shout still*. Three students simply ignored the word *still* and only provided *was shouting* (and one student *shouted*) in slot 16; the same three students behaved similarly in slot 25 and used *was talking* in contrast to 247 students writing *was still talking*. In slot 108, two of them used the word *still* and one did not provide an answer. Interestingly, one student wrote *was still and shouted* in slot 16, which actually sounds quite contradictory in the context.

[97] Unlike in some previous sections, double forms are included in the discussion in this section. This is done because it is occasionally precisely because of these forms that the interesting phenomenon arises.

Unusual word order was not particularly frequent and students did not seem consistent with unusual word order. Thus, they mainly appeared to be occasional slips, not systematic patterns. However, while four students ignored the word *still* in slot 16 and two ignored *obviously* in slot 17, all students used the word *normally* in the regular fashion in slot 42, except for one student, who either ignored the adverbs in each of the slots or did not provide an answer at all in such cases. In slot 39, for example, she[98] ignored the word *never* but retained the negativity and provided the form *didn't like*. Only one student produced *like never*; other students either had the word *never* first or in the middle of the verb phrase. Three other students also just used a form with either *not* or *not ever*. Interestingly, no student used a double negative, such as *didn't never*. Actually, double negatives did not occur at all in the study: this seems to be something that the students master very well.

One person dropped the word *already* in slot 70, and most students (291) placed the word in the middle of the verb phrase when they used two-verb forms; only four students placed the word *already* before the two verbs. One student used three verbs and placed the adverb after the second verb, i.e. *has been already leave*. Of the students who only used one verb, 17 placed the adverb before the verb and 5 after the verb. One student changed the word *even* to *never* in slot 100, but all the other students provided *even* in their answers (except for the five who provided no answer). Again, the word *even* was mostly provided in the middle of the negative verb phrase. In slot 108, only three students placed the word *still* after the verb phrase. In contrast, *takes normally* or *took normally* was provided by 41 students, as opposed to 240 students providing *normally takes* or *normally took* in slot 42. This implies that it may actually be easier for students to place the adverb correctly when there are two verbs, as they seemed to make fewer errors in these cases. Another possible explanation is that students were not equally familiar with the word *normally* and thus had more difficulty in deciding where to place it.

In addition to dropping adverbs, some students dropped prepositions. There were two slots where a preposition was to be used in the verb phrase, and in slot 108, as many as 68 students ignored the preposition *at* in their response, while only two ignored *out* in slot 101. This may be because the required verb form in slot 108 was more complicated than in slot 101; furthermore, perhaps students were more familiar with *find out* than with *shout at*. It may also be that students did not even realise they were supposed to include the preposition, especially in slot 108.

[98] Gender-specific pronouns are used throughout in this chapter, because using 'they' would be clumsy.

Interestingly, some students seemed confused with the verbs *go* and *get* and mixed their forms. In slot 71 (**... before I ___ (get) there**), which was the only slot in which a form of *get* was supposed to be used, 17 students provided *went*. Fortunately, this change still made the sentence understandable and retained the intended meaning. Furthermore, students were expected to use a form of *go* in eight different slots, but some used a form of *get* in each of them. Thirteen students provided *got* in slot 20, and two provided *get*. One student also provided *had got* and another one *has got*. In slots 21 and 44, one student used a form with *get*. Three students used a form of *get* in slot 69, and 16 students did so in slot 88; of these, two also did so in slot 89, while the others properly used a form of *go*. Six students used a form of *get* in slot 97 and two in slot 100. A handful of students did this more than once, but no-one was entirely systematic in using *get* instead of *go*. Of the eight slots, Student 290 used *get* instead of *go* six times, and Student 126 five times. Two students did so three times and six student two times, while 36 students only erred once. Only Student 45 did this both from *go* to *get* and *get* to *go*.

In addition to the confusion between *go* and *get*, there were other, apparently more deliberate changes in the verbs. Some changes were fairly small: in slot 5, four students used a form of the phrase *be in love with* instead of just *love*; one student changed the plain *call* to the phrase *make a call* in slot 12; one student preferred *have a call* to just *call* in slot 59 and one student used *took a look* instead of *was looking* or *looked* in slot 91. One student changed from *want* to *wish* in slot 7, and two students offered a form of *kill* and one student a form of *slaughter* instead of *murder* in slot 6. Again, one student provided *went* instead of *came* in slot 24, and one student favoured *saw* as opposed to *met* in slots 92 and 107 (a different student in each slot). A few more students changed the verb in slot 28: instead of using the required *phone*, nine students offered a form of *call*. Seven students added either *me* or *to me* after either *phoned* or *called*, perhaps thinking that it should be clarified who phoned or called whom. Two students changed from *walk* to *go* in slot 55, and in slot 109, one student decided to write *is lying* instead of *is telling*, although the next word after the slot was *lies*. In the last slot (110), one student used *do* instead of *make*. The teachers' reactions to some of the changes are discussed in Chapter 9.2.2.

The four forms of *spoil* and *spell* in slot 34 were probably not intentional verb changes but spelling errors, since their meaning is quite different from the intended *spill*. The same probably applies to *sawed* in slot 47, where *saw* (as the past form of *see*) was the expected form; the same happens in slot 81 with *sawing* (7 students) instead of *seeing* and *had saw*, *had sawn*, *has been saw*, *has saw*, *saws* and *was saw* (altogether 30 students) instead of *had seen* in slot 95 (**I asked her if she ___ (see) the dog, but she said ...**). For some reason, one student used the verb *save*

instead of *see* in the same slot. Perhaps he intended to provide some form of *saw* but became confused. Changing the verb was not the typical behaviour for any student; only nine students changed the verb twice, the others were single instances.

At times, some students either added or dropped words in the slots. Dropping the adverbs or prepositions was discussed earlier in this section, but some students also inserted prepositions, pronouns or other words in the slots. In addition to the cases in slots 5 and 28 that were discussed above, one student wrote *poured it* in slot 35 and another *never like for* in slot 39. One student wrote *took me* in slot 43. In slot 60, four students added the word *there* after *was* or *were,* and in slot 63, one student continued the verb *seemed* with *to be*; seven students did so in slot 94. One student in slot 71 wrote *got (in)*, and another used *walked up* in slot 93. Four students wanted to use the word *seen* instead of just a form of the word *have* in slot 96, for example *hadn't seen*, and two students also used the word *it* in addition to *seen* to refer to the dog, as in *had not seen it* / *hadn't seen it*. Some students provided the subject *I* with the verb form, for example in slots 5 and 6, although it had already been provided in the test, and one student repeated the preposition that was provided in the test in *was looking for* in slot 58. Similarly, although the word *me* was provided after slot 74, 13 students also included it in their response, for example in *didn't let me*. Finally, in slot 77, the preposition *about* was given in the text but repeated in the response by eight students.

Eight students noticed a missing article in the test after slot 42 (**It ___ (normally/take) quarter of an hour to walk to the village**) and provided it with the verb form, with, for example, *normally takes a* and three other forms. Whether other students also noticed the lack of the article remains unclear, but at least these eight added it in their responses. A number of students felt the need to insert an object in the verb phrase *find out* in slot 101: here, 26 students used the word '*it'*, as in, for example, *found it out* or *found out it*, while one student used '*that'* in *had found that out*. One student added *that* after *means* in slot 102, and one used *by* instead of the preposition *at* in slot 108. In slot 109, one student repeated the word *lies*, which was provided in the test. Again, in most cases, students added a word only once in the test, but ten students did so twice and two students, Student 180 as well as Student 245, three times.

Students occasionally used the emphatic *do* or *did* as well, although it was not actually required in any of the slots[99]. It was used 56 times in 35 slots, which is 0.5% of the forms used in these

[99] This discussion does not include forms that consist of *did* and *–ing*, as in *did planing*, as this is a form that is difficult to interpret.

slots. These were mainly provided by only one or two students per slot, but there was one instance (slot 47: **I ___ (see) Gerald, that's Dad's business partner**) where six students provided such a form and three instances where three students provided such a form. There were three students with *did notice* in slot 13, *did hear* or *did heard* in slot 26 and *did leave* in slot 103. Five students decided to use *did see* and one student *did saw*[100] in slot 47. It may be that these students wanted to emphasise the contrast between no-one seeing Lucy in slot 45 (**No one ___ (see) me when...**) but her seeing Gerald in slot 47 (**I ___ (see) Gerald, that's Dad's business partner**) – and perhaps even the contrast with him not seeing her in slot 49 (**He ___ (not/see) me, though, because it was dark outside**). Another possibility is that they anticipated the potential of someone denying her having seen Gerald, which may have driven the students to use the emphatic form; one further possibility is, of course, that these students did not really know how to use the emphatic *do*. Twenty-three students used emphatic forms, but most of them only once. Students 98, 120, 128, 159, 162, 198 and 211 were more productive with the emphatic form, with the highest number being 10 slots for Student 128, who can already be considered to overuse them.

8.7 English Majors compared with other students

As we saw above in Chapter 7, the English Majors made fewer errors on average than the other groups. Obviously, they were also, *a priori*, assumed to be the best-achieving students. The aim of this section is to explore in what ways they behaved differently from the other groups. For this reason, the discussion about inter-group variation focuses on the distinction between English Majors compared with all the other groups, because this helps to identify the ways in which they stand out.

As we saw in Section 7.1.1, there were 107 slots to be filled in in the test. Of these, all the 35 English Majors gave an identical and correct answer in 25 slots. In 24 of these, the correct answer was the simple past form in the affirmative and in one, the answer was an *–ing* form following a preposition. In comparison, for example the Grammar group only once uniformly provided only one form per slot, *walked* in slot 55.

[100] However, whether this intended to be an emphatic form remains unclear.

Although most English Majors were skilled at distinguishing between simple and progressive forms, some English Majors supplied the simple form when the progressive was expected (see Section 8.1.2). The students using the simple instead of the progressive more frequently were not always systematic: for example, while 10 English Major students used the simple in slots 32 (**Mum ___ (watch) some stupid film after dinner, so ...**) and 35 (**... while I ___ (pour) it**), three of them were not the same students, but seven were. The degree to which students used the simple instead of the progressive depended on the slot: for some slots, a third of the English Majors ignored the progressive, while in some other slots, only a few students did so. For example, 24 students provided the progressive *was watching* in slot 10 (**I ___ (watch) a rather exciting film on television, so ...**) and 11 students the simple *watched*, but in slot 25 (**... while the doctor ___ (still/talk) to Trevor)**, only two students provided *still talked*, while 33 provided *was still talking*. In comparison, the Grammar group was split in half with using the simple instead of the progressive in slot 10, with 18 students providing *was watching* and 22 providing *watched* (and two students providing yet something else), and in slot 25, seven students used *still talked* and 28 used *was still talking* (seven students provided other responses). It seems, then, that with the word *still*, students were more likely to think of using the progressive than without it. The use of the progressive is discussed more extensively in Section 8.1.2.

Another problem for some English Major students was the use of the past tense where a past perfect form would be required, as in slots 4 and 53. Of the 35 English Majors, 13 students provided *was murdered* and 18 provided *had been murdered* in slot 4 (**It was soon clear that he ___ (murder)**), and 22 students provided *had told* and 13 *told* in slot 53 (**he __ (tell) me he ...**). However, there were some other slots where nearly everyone provided the past perfect, such as slot 95 (**I asked her if she ___ (see) the dog, but ...**) with *had seen* by 33 students, or slot 70 (**Lucy ___ (already/leave) the room before I ...**), with *had already left* by 34 students. While the word *already* may have helped students in slot 70, no such word was provided in slot 95; nevertheless, most English Majors were very good at finding the right tense for this slot. When we compare the English Majors with the other student groups, we notice that the other students used the past perfect less often, with only about 21% of the students providing *had been murdered* in slot 4 but 43% providing the past perfect form in slot 53. The word *already* in slot 70 seems to have helped them, too, since the score of the past perfect form was almost 70% for slot 70. Actually, the pattern for using the past perfect was very similar for the groups; when a greater proportion of English Majors used the past perfect, so did the other students. The use of the past perfect is discussed further in Section 8.1.3.

In Section 8.5, we discovered that students tended to use contracted negative forms, with well over half of the students favouring them. However, there were proportionally more students who did not use contractions in the English group. The distribution varied somewhat from slot to slot, but generally the students were split in half. For example, in slot 76 the difference was one student: 17 spelled *didn't know* and 16 spelled the form *did not know*; two students provided *doesn't know*. It is possible that English Majors had been instructed to avoid contracted forms in their studies or that they were more aware of stylistic differences between various genres.

At times, when students provided a number of different forms, teachers also seemed undecided on the best form to use. This was often reflected in English Majors' responses so that their distribution of such competing forms was more even than with the other students and more closely followed the teachers' opinions. For example, in slot 14 (**... because we ___ (expect) him to come earlier**), where the vast majority of all students (182) provided *expected* and some also *were expecting* (65) and *had expected* (44), the English Majors more evenly used the three forms (13, 9 and 10 students, respectively). Teachers were somewhat torn between these choices, with the majority supporting *had expected* as well as *had been expecting*, which only three students in the entire population produced (one of them was an English Major).

English Majors seemed very proficient in placing adverbs in the correct place in the verb phrase, while other students were more insecure about the placement of adverbs such as *obviously* and *still*. While English Majors generally were good at using irregular verbs, three students experienced difficulty with the past participle of *go*, using *had went* and *have went* in slot 20. While *have went* was the only instance in the entire population, 10 students from other groups also used *had went* in this slot. Three English Majors also used a regular past tense ending for the word *hear*, producing *heared* in slots 23 and 26. This was also witnessed in students from other groups, with 27 and 26 students, respectively. Some attempts at making a regular verb irregular also occurred: One English Major student thought the past simple of *remember* is *rememberd*, and one student created an unusual past form from *creep*, providing *crep*[101]. Students from other groups were also productive in this slot (slot 37: **... so I ___ out by the back door**), offering forms such as *creapt*, *crop*, *creped* and *creaped*. While eight English Majors thought the past simple of *creep* was *creeped*, the majority of students in other groups were not aware of the fact that this verb is irregular, either: 171 students provided *creeped*, 117 provided *crept*.

[101] The lack of the letter *t* cannot be explained through assimilation, for the following word is '*out*'.

There were not many students who used *would not let* / *wouldn't let* in slot 74 (**... but he ___ (not/let) me**), but of the ones who did, English Majors constituted a large percentage. Ten English Majors used such forms, while the total for all students was 21 (plus one with *would not let me*). This may be a usage that is not commonly discussed at school, and only those more devoted to English would be aware of this structure being used in the sense of refusal or criticism. *English Grammar in Use* explains that the structure "[s]omebody **wouldn't do** something" means that "he/she refused to do it" (Murphy 2004, 72, emphasis original). Another grammar book, *Advanced Grammar in Use*, says that *would* can be used to "criticise people's characteristic behaviour or habits", often to suggest "that criticisms have been made before but ignored", but it should not be used for situations where "a single event happened at a given past time" (Hewings 2005, 32). Actually, the context does not reveal whether the murdered man had the habit of not listening to his doctor or whether this only happened once in this occasion, but it seems that at least some students either made the interpretation of this being his habit from the context or else understood it in the sense of refusal. Teachers also seemed to approve of making these assumptions, since they rated the forms *would not let* and *wouldn't let* mainly with the best label, although most of them also accepted *did not let* / *did not let* as the best choice.

In some cases, English Majors also encountered spelling problems, although their prevalence was much smaller than with other groups. These spelling problems were limited to three types: the doubling of consonants before -*ing* or -(*e*)*d*, the formation of the past tense with -*ed* or just -*d*, and the retaining or addition of an *e* before *–ing*. One English Major student doubled the consonant in the progressive form of *look* and *shout*, thus systematically producing a form including *shoutting*[102] four times and *shoutt* once, but she produced *lookking* only once and did not double the *k* in the other slot where the form occurred but spelled *looking*. The same student's past simple form of *plan* became *planed* and of *stop* became *stoped*. Two other students created similar deviant forms, Students 11 and 12 with *stoped* and Students 11 and 14 with *had planed*. Student 11 further added an *e* before *–ing* with *was still beeing shouted at* and Students 21, 24 and 35 kept the *e* in *argueing*. In comparison, students from the other groups displayed a variety of spelling errors in addition to the ones mentioned here, from misspelling the word that was given to creating a number of innovative forms with unusual spelling. These include *haventnt answered* (slot 51), *walhed* (slot 10), *planted*[103] (slot 61) and *to exlpain* (slot 72).

[102] Another English Major once spells *shoutting*, but does not double the *t* in the other slots.
[103] This is a deviant form because the verb you were supposed to use was *plan*.

Of the slots towards the end of the test where several students faced difficulty, English Majors did better than the rest of the groups, although some also made errors. A third of the English Majors, 10 students, accurately provided *was still being shouted at* for slot 108 (**She ___ (still/shout at) by her father at nine-fifteen**), while in the entire population only 28 students (9%) found this solution. In the last slot (slot 110: **I think it's time we ___ (make) an arrest**), 15 out of 35 English Majors used *made* while 20 used *make*; in the entire population, a significant number of students, 200, provided *make* and only 49 *made*. In one further slot (slot 79: **... it was no use ___ (argue) with him so ...**), the majority of English Majors, 23 students, provided *arguing* and 8 students provided *to argue*, while in the whole population, 101 students provided *arguing* and 158 thought the form is *to argue*. Thus, English Majors seemed to be more familiar with grammar in these fixed phrases.

8.8 Summary

Finnish students displayed wide variation in their responses to the test. Among the reasons for the variation were the use of an unconventional tense or aspect, spelling errors and difficulty with word order. Some students had also either added extra words, for example prepositions, to their responses or ignored the words, for example adverbs, that they were supposed to use. Some had even decided to substitute the verb with a different verb, although usually with a similar meaning.

While most students were successful at providing the past simple where it was expected, the other forms were more challenging for them. Even some English Majors had problems with identifying cases where the progressive form was needed instead of the simple and partly also with the past perfect. Overall, however, they tended to be better at spelling and were significantly more proficient in the more challenging forms in particular, and barely made any errors with simpler forms, such as the past tense.

Some Finnish students found it difficult to remember the changes to verb forms caused by reported speech, perhaps because similar changes do not occur in Finnish. Although most Finns were good at remembering irregular inflection, the verb *creep* proved to be more difficult, and not many used *spilt*. The most difficult structure turned out to be the combination of the progressive past in the passive, combined with a preposition and an adverb. Furthermore, some

students were not familiar with particular fixed phrases in English. Some students had difficulty in placing the adverb in the verb phrase, and a few overused the emphatic *do*.

While some students were inconsistent in their responses, the majority held an established system of either using contracted or uncontracted verb forms. Contracted forms were common in the negative and rare in the affirmative, while English Majors tended not to use contracted forms as frequently as the other groups. Some slots remained ambiguous and could justifiably be interpreted in several ways.

9 Variation in teachers' responses

After a discussion of the students' skills, it is necessary to explore the range of variation in teachers' opinions regarding the acceptability of the various forms the students had provided. As was discussed above, the degree of consensus varied from form to form, and a number of slots were controversial. Indeed, teachers of English were not always unanimous in determining what the "correct" form was in a particular slot. In several cases, the same form may be the best possible in one teacher's opinion and completely unacceptable in the eyes of another teacher. This chapter delves deeper into the nature of disagreement in the teachers' assessment. After a short overview below, the extent of agreement and disagreement is discussed in Sections 9.1 and 9.2, where the first one (Section 9.1) focuses on cases where only one teacher disagrees with the remaining 12 teachers, and the later (Section 9.2) discusses more extensive disagreement among teachers. Section 9.3 comments on the interplay between teachers' self-reported approach to grammar and the nature of the responses they provided for this study. The teachers' background information is available in Table 7 in Section 6.3. Throughout this chapter, it is, again, useful to refer to Appendix 2, where the test is available. However, wherever practical, the immediate slot context is provided[104]. The full list of the student responses and the teachers' assessment is provided in Appendix 6.

In total, there were 1,522 verb forms in the study (available in Appendix 4) for the teachers to evaluate[105]. This consisted of 5 to 47 verb forms per slot; there were 107 slots. In total, the teachers provided 19,786 ratings: 2,258 forms (11%) were rated as 1 (the best), 867 forms (4%) as 2 (acceptable), 439 forms (2%) as 3 (questionable) and 16,151 forms (82%) as 4 (inappropriate). Missing or ambiguous ratings have been coded with ? (a question mark); there were 71 such instances. This means that teachers found approximately 16% of the verb forms appropriate for the context and dismissed 84% of them. The distribution is provided in Table 36, while the rating scale is explained in Section 6.3 and is also available in Appendix 3.

[104] As in Chapters 7 and 8, I have systematically dropped the slot numbers from the slot quotations.
[105] Section 11.2 explains why the teachers worked on a slightly different list compared to the one students ultimately provided.

Table 36. The distribution of teacher ratings

	1 = the best	**2 = acceptable**	**3 = questionable**	**4 = inappropriate**	**? = no response or ambiguous**	**Total**
Number of ratings	2258	867	439	16151	71	19786
% of the ratings	11.41	4.38	2.22	81.63	0.36	100

Of all the verb forms, all the 13 teachers who participated in the study unanimously indicated 64 verb forms as the best choices for their particular slots[106]. Although these teachers were allowed to nominate several best choices and they did so, there was never more than one alternative per slot that was accepted as the best by all the teachers, and with 43 slots, there was no consensus on what the best choice for that slot would be. However, the teachers were unanimous in rating 825 verb forms as inappropriate. Furthermore, of the 71 cases where a teacher either did not provide a rating or gave an ambiguous one, 36 occurred in forms where everyone else marked them as 4 (inappropriate), and one in a form which everyone else rated as 1 (the best).

The teachers showed the greatest consensus with slot 30 (**I ___ (tell) her ...)**, where there were only five alternatives to rate. Of these, teachers unanimously rated *told* as the best alternative, while *telled*, *told to* and *toll* were all unanimously rated inappropriate and the only form with any disagreement was the form *tell*, which one teacher found acceptable and the remaining 12 teachers marked as inappropriate. Quite frequently, the more student variation there was, the greater the degree of teacher disagreement was as well.

9.1 Single teacher disagreement

In addition to the 889 verb forms that all teachers agreed on, there were 203 verb forms where one teacher disagreed with the remaining 12 teachers and 174 verb forms where two teachers disagreed with the remaining 11. Thus, there also remains a significant number of verb forms with more extensive levels of disagreement. Some of these only occurred within the range from the best to acceptable, but many also range from the best to entirely inappropriate. Some teachers were more willing to consider alternative forms than others, which is not surprising.

[106] One further form could perhaps be considered to belong to this category, because there was also one form that 12 teachers rated as being the best and one teacher had not rated at all.

Individual variation is summarised in Table 37 for the cases where only one or two teachers diverged from the rest of the teachers. The cases where only one teacher disagreed are discussed in this section, and those where two or more disagreed are discussed in Section 9.2.

Table 37. The number of times teachers disagree with the others

Teacher number (nationality)	**The only one to disagree when 12 teachers rate the form with 4**	**One of two[107] teachers to disagree when 11 teachers rate the form with 4**	**The only one to disagree when 12 teachers rate the form with 1**	**One of two teachers to disagree when 11 teachers rate the form with 1**	**total**
1 (Finnish)	2	3	8	5	18
2 (Finnish)	7	13	0	1	21
3 (Finnish)	28	33	1	1	63
4 (Finnish)	8	9	1	1	19
5 (Finnish)	15	10	2	2	29
6 (Finnish)	37	19	0	1	57
7 (British)	0	2	2	3	7
8 (British)	1	5	1	2	9
9 (British)	36	22	0	1	59
10 (American)	2	0	2	1	5
11 (American)	6	4	1	2	13
12 (American)	34	27	6	4	71
13 (American)	2	1	1	2	6
total	178	148	25	26	377

In this study, Teachers 3, 6, 9 and 12 were the most likely to mark a form with 1, 2 or 3 when the rest of the teachers found the form inappropriate (4). Furthermore, Teachers 6 and 9 used the 'acceptable' category particularly frequently, while Teachers 3, 5 and 12 were more eager to use the 'questionable' category. Teacher 3 tended to use the label 'questionable' to some forms that are grammatical but not the suitable tense or aspect in the context, or when the student had added extra words to the response, such as *walked up* instead of just *walked* in slot 93. She[108] also found questionable some forms where the verb had been changed, as in *did not even get* instead of *did not even go* in slot 100. Interestingly, she only began to use the 'questionable' category from slot 42 onwards; in contrast, Teacher 5 mainly used the 'questionable' category before slot 35, using it only once after that. A similar phenomenon occurred with teacher 12, who frequently found forms questionable at the beginning but only had two questionable instances after slot 42. Teachers 5 and 12 tended to label 'questionable' forms that are grammatical but not the standard choice in the context. Teacher 12 also found

[107] There were also ten cases where one teacher had not rated a form at all, 11 had rated them as 4 and one teacher as 1, 2 or 3. These cases are included in the table in this column despite the fact that it remains unclear whether there is one or two teachers who disagree.

[108] Gender-specific pronouns are used throughout in this chapter, because using 'they' would be clumsy.

'questionable' forms where the intended affirmative verb had been made negative, as in *did not expect* in slot 14 (**... because we ___ (expect) him to come earlier**).

Teachers 1 and 12, on the other hand, were the most likely to mark a form with 2, 3 or 4 when the others found it the best alternative (1). Teacher 1 occasionally had a tendency to disapprove of the use of contracted forms in cases where he accepted the uncontracted form. At times, he distinguished between the best and acceptable, but in some slots, he found the contracted form inappropriate; other teachers tended not to distinguish between the contracted and uncontracted forms as systematically but did so occasionally. Teacher 12 occasionally disagreed with the tense or whether the progressive form should be used in a particular slot. Teachers 7, 8, 10 and 13 were not likely to deviate from the consensus.

There were nine instances where one teacher disagreed with the others most radically so that while 12 teachers rated the form as being the best, one teacher felt it was inappropriate. Teacher 1 thought the best answer in slot 60 (**I ___ (be) rather later than ...)** was not *was*, as everyone else thought, but *would be*. He also found *weren't* in slot 82 (**We ___ (not/be) really friends**) inappropriate and only accepted *were not*. Teacher 3 thought *is telling* in slot 109 (**Someone ___ (tell) lies**) was inappropriate and suggested *has been telling* as the best form. Teacher 4 disagreeed with *shouting* in slot 19 (**Then Trevor stopped ___ (shout)**) and favoured *the shouting* instead. Teacher 5 thought the response to slot 83 (**Yes, my house ___ (be) just round the corner from the Sterns'**) was not *is* but rather *was*. Teacher 7 only accepted *didn't see*, not *did not see*, in slot 49 (**He ___ (not/see) me, though, because it was dark outside**). Teacher 8 did not find any of the forms in slot 7 (**After dinner last night he said he ___ (want) to check some business papers in his study**) the best[109], while the others thought it was *wanted*. In slot 27 (**... during a quiet few seconds when Trevor ___ (not/shout)**), Teacher 10 only accepted *wasn't shouting* and not *was not shouting*, and in slot 4 (**It was soon clear that he ___ (murder)**), the same teacher supported *had been murded* instead of *had been murdered* as the best alternative. Although many of these single instances were perhaps unintentional slips, it is likely that at least Teachers 1, 3 and 5 were convinced that the form they supported truly suited the context better. One teacher (Teacher 12) also found 'questionable' four forms which the rest of the teachers accepted as the best. These were *didn't murder* in 6 (**But I ___ (not/murder) him, either**), *had told* in 53 (**... he ___ (tell) me he ...**), *was looking* in 91 (**I ___ (look) for him when ...**) and *was walking* in 93 (**She ___ (walk) up the road towards their**

[109] This is probably not intentional but a random slip.

house). Instead, he supported *did not murder* in 6, *told* in 53, *looked* in 91 and *walked* in 93. Thus, in two instances this teacher thought that instead of the progressive, the past simple should be used.

There were also 15 instances where the disagreement was equally radical, but in the opposite direction: in these instances, 12 teachers rated a form as being inappropriate, but one teacher found it the best. Teacher 2 was the only one to accept *have gone* in slot 44 (**I can't prove I ___ (go) to the village**), while Teacher 3 accepted *werent* in slot 82 (We **___ (not/be) really friends**). Teacher 4 accepted *did not* in 4 (**It was soon clear that he ___ (murder)**), *haven't ever liked* in slot 39 (**I ___ (never/like) Mum or Dad to be around when ...**), *be* in slot 75 (**He said I ___ (be) an ignorant country doctor who ...**) and *did not* in slot 106 (**Lucy ___ (not/walk) to the village and back, if ...**), and Teacher 5 accepted *to shout* in slot 19 (**Then Trevor stopped ___ (shout)**). Teacher 6 found the form *expect* the best in slot 14 (**... because we ___ (expect) him to come earlier**) and *did tell* in slot 109 (**Someone ___ (tell) lies**). Teacher 9 accepted the form *go* in slot 88 (**I ___ (go) up on the hills, away from the village**), and Teacher 10 accepted *had been murded* in slot 4 (**It was soon clear that he ___ (murder)**) and *like never* in slot 39 (**I ___ (never/like) Mum or Dad to be around when ...**). Finally, Teacher 12 accepted *didn't notice* in slot 13 (**I ___ (notice) the time because ...**) and *didn't even* in slot 100 (**She ___ (not/even/go) into the study**), and Teacher 13 accepted the form *'d planed* in slot 61 (**... rather later than I ___ (plan) to be because ...**). Four teachers (1, 7, 8 and 11) never indicated something as the best without the others also doing so.

At times, one teacher felt that a given form was perhaps not the best but still acceptable, while all the others considered it inappropriate. This was the case with exactly 100 forms. The forms that only one teacher accepted were often alternative tense forms[110], changes to the verb to be used, such as *saw* instead of *seemed* in slot 64 (**... she ___ (seem) rather embarrassed**), and occasionally slight spelling errors, as in *didn't wont* or *didnt want* in slot 22, where the expected form was *didn't want* / *did not want*. Teachers 6 and 9 were particularly prominent in this category, together being responsible for ⅔ of the cases where one teacher marked a form as acceptable while the rest did not approve of it at all. The two teachers were, however, dissimilar in what they accepted. Teacher 9 accepted a vast variety of spelling errors, for example *shoutting* in slot 19 or *dicided* in slot 38. At times, she also accepted subject-verb disagreement, as in *was obviously having* in slot 17 (**He and Lucy ___ (obviously/have) a serious row**).

[110] Some of these occurred in places where various alternative interpretations can readily be created, as discussed in Section 8.2.

Teacher 6 did not tend to accept spelling errors but often accepted the present tense where the rest required a past tense, as in *seems* in slot 94 (**She ___ (seem) rather upset**) and in *tell* in 30 (**I ___ (tell) her ...)**. Perhaps she considered these forms instances of the so-called historic present, where events located in the past are narrated in the present tense "to produce a more vivid description, as if the events were being enacted at the time of speech" (Biber et al. 1999, 454). This occasionally occurs in fiction, in conversational narratives in particular.

In conclusion, Teachers 7 and 8 (both British) as well as 10 and 13 (both American) were not as likely as the others to be the only ones to seek for alternative ways of interpreting the slots. These teachers could perhaps be characterised as conforming to standard forms. In contrast, Teachers 3 and 6 (both Finnish), 9 (British) and 12 (American) were more likely to accept other ways of understanding particular contexts, and they could be characterised as occasionally being unconventional ("adventurous") in their assessment. The middle ground is the position for Teachers 1, 2, 4, 5 (all Finns) and 11 (American). Interestingly, the Finnish teachers were more likely than most of the native speakers to deviate from the others in cases when only one teacher differed from the rest. The next section discusses more extensive teacher disagreement, where two or more teachers provided a different understanding of a slot compared to the others.

9.2 More extensive disagreement

When several teachers disagreed with one another and when some of these teachers found the form the best while others found the form inappropriate, the picture becomes fairly complicated, and the interpretation of the situation at hand in the story plays a greater role. Some teachers accepted the progressive when the simple would be the standard, and vice versa. For some teachers, there was a difference in the acceptance rates for contracted and uncontracted forms. For this reason, both forms are given in this section whenever the discussion affects the two.

Some previous research (see Chapters 2 and 3) has argued that native and non-native teachers rate students' responses differently. Furthermore, there can be some difference between the acceptance of particular forms depending on whether the raters have acquired English in the United Kingdom or in the United States (see Chapter 2). To address this, disagreement that can be traced to a variety or linguistic background, at least partly, as well as disagreement related to specific phenomena, is discussed in Section 9.2.1, while Section 9.2.2 is devoted to more individual variation and disagreement.

9.2.1 Group-level variation

This section addresses variation caused by two or more teachers in features where some distinction can be created based on whether the teachers are Finnish, British or American. When two to three teachers rated a form with 2 or 3 while the rest found the form inappropriate, the form typically included either an unusual interpretation of time, as in *don't love* for slot 5 (**I ___ (not/love) my husband, he was a cold and selfish man)**; the emphatic *do* (see e.g. Biber et al. 1999, 433-435) in slots where there was barely any need to use emphasis, as in *did take* in slot 18 (**So I ___ (take) the doctor into the sitting-room for a moment**); or unusual forms in reported speech, as in *have decided* in slot 31 (**... I ___ (decide) to leave Trevor)**. Some teachers also allowed inserting additional words (e.g. inserting the word *it* in *poured it* in slot 35) or ignoring words that were intended to be used (e.g. ignoring the word *obviously* in slot 17 (**He and Lucy ___ (obviously/have) a serious row)** with *seemed to have*). Some teachers also allowed unusual word order (e.g. *took normally* in slot 42 (**It ___ (normally/take) quarter of an hour to walk to the village**) or being innovative with the response (e.g. *took a look* in slot 91 (**I ___ (look) for him when ...**)).

At times, Finnish teachers seemed to be more willing than British or American teachers to accept the simple form where the progressive is the standard. For example, all of the British and American teachers found *talked* in slot 50 (**He ___ (talk) on the phone, I think)** inappropriate, but two Finns found it acceptable and two questionable; two Finns also found it inappropriate. Half of the Finns found *talked* in slot 77 (**... what he ___ (talk) about**) the best form, while none of the American or British teachers did so. However, a different profile appears in slot 32 (**Mum ___ (watch) some stupid film after dinner, so she ...)**: two Brits and two Americans found *watched* acceptable or the best form, while only one Finn found the form acceptable, and in slot 35 (**... a few drops of tea on his desk while I ___ (pour) it**), three Finns, one Brit and two Americans found *poured* acceptable or the best.

British and American teachers were somewhat more tempted to allow the unusual progressive *was not answering* / *wasn't answering* in slot 51 (**Alan ___ (answer) the phone)**, while Finns were more tolerant with the progressive *was reaching* in slot 57 (**... just before I ___ (reach) our house**). American teachers were also willing to accept *are not* / *aren't* in slot 82 (We ___ **(not/be) really friends**), while the Finnish and British teachers were not. In the same slot, some Americans and some Finns but no Brits accepted *had not been* / *hadn't been*. On the other hand,

all the British teachers accepted *had come* in slot 24 (**I think Lucy ___ (come) into the house while ...**), while American and Finnish teachers were not fully convinced: half of them accepted the form.

Both British and American teachers found *explaining* in slot 72 (**I tried ___ (explain) to Trevor why ...**) appropriate, while Finns were more hesitant. In slot 74 (**... but he ___ (not/let) me**), again, two of the British and one American teacher accepted both *did not let* / *didn't let* and *would not let* / *wouldn't let*, while generally the Americans preferred *would not let* / *wouldn't let* and Finns *did not let* / *didn't let*. One Brit did not accept *did not let* / *didn't let* and one Finn did not accept *would not let* / *wouldn't let*. Furthermore, two Finns found the contracted form *didn't let* inferior to *did not let*: one Finn marked the contracted form as acceptable while the uncontracted was the best, and another found the contracted form inappropriate but the uncontracted form the best.

Two American teachers found *have been living* inappropriate in slot 84 (**I ___ (live) here for two years now**), but all the British, two of the Americans and the majority of Finns found the form acceptable or the best. Three Americans preferred *'ve lived* in slot 84, but some of the Finns, one Brit and one American found that form inappropriate. All of the British teachers found *were talking* in slot 29 (**... and we ___ (talk) for ages**) the best, while most of the Americans and Finns found it inappropriate or questionable.

As we can see, it remains difficult to trace a particular pattern of variation in the assessments discussed above as none of the phenomena is systematic, and given that there were only 13 teachers, perhaps it is not feasible to generalise on the basis of the variety when there were so few representatives per variety in the study. However, Finnish teachers were perhaps a little more lenient with unusual spelling and unusual tense than British and American teachers. For example, in slot 104 (**Anyway, Dorothy Stern told her sister she ___ (leave) her husband**), several Finns accepted *'d leave* and *'ll leave*, while almost all British and American teachers found them inappropriate, apart from one American accepting *'d leave*. Furthermore, some Finnish teachers more readily accepted changing the verb, adding words and using the emphatic *do*, but individual differences were noticeable. In slot 71 (**... before I ___ (get) there**), four out of six Finns accepted adding the word *in*, i.e. writing *got in*, but none of the Americans did; only one British teacher did so as well. However, in slot 21 (**I think he wanted to persuade Trevor ___ (go) to the hospital for some tests, but ...**), it was the Americans in particular who accepted *to go to,* although the latter *to* is not needed. The Finnish teachers were somewhat

more tolerant of keeping the present tense in reported speech, with five Finnish teachers accepting *does not know* in slot 76 (**... who ___ (not/know) what he ...)**, while none of the British teachers and half of the Americans did so.

Finnish teachers seemed to be somewhat more willing to accept emphatic past forms when they were used in contexts where the emphatic form was not expected, as in *did hear* in slot 26 (**I ___ (hear) the front door bang during a quiet few seconds when ...**), which two Finns found acceptable and two questionable, while all of the Americans or British teachers found such forms inappropriate. The form *did leave* in slot 103 (**He ___ (leave) the Sterns' hours before half past nine)** is another example: two Finns found it the best form, one found it acceptable and one questionable, while all of the British and Americans found the form inappropriate. This may reflect a difference in how the teachers have learned the language: perhaps the Finnish teachers, being more reliant on the application of the rules they have acquired, did not fully apply them in the manner intended and were perhaps excessively lenient with emphatic forms. On the other hand, the nature of input and the types of exposure the various groups have received may also account for the difference. Furthermore, different users of the language may be different in how likely they are to accept emphasis or creativity with language.

In slot 99 (**It ___ (not/be) his wife**), some Finns and some British teachers were willing to accept *can not be* and *wouldn't be*, but all of the Americans found the forms inappropriate. The Americans were also more hesitant with *can not have been*, although most of the others found this form acceptable. The Americans were happier with the forms *cannot be* and *can't be*. However, the range was wide: in this slot, Teacher 6, a Finn, and Teacher 10, an American, indicated a dozen 'best' forms, while Teacher 1 only labelled one and Teachers 12 and 13 two forms as being the best. Americans also disfavoured *couldn't see* in slot 49 (**He ___ (not/see) me, though, because it was dark outside**), while the British and Finnish teachers (except for one) found it either the best or an acceptable form. In slot 100 (**She ___ (not/even/go) into the study**), again, all the British found *had not even gone* / *hadn't even gone* inappropriate, while half of the Finns and Americans found the forms acceptable in the context. While all the Americans and two of the British marked *had* in slot 8 (**He ___ (have) a meeting with Gerald, his business partner, the next morning**) with the best label, one Brit and four of the Finns were not convinced that this form was suitable at all. This is perhaps because although the meeting was planned, it did not ever take place because the person was murdered. However, this being an instance of reported speech, some teachers found the use of this form entirely

acceptable. Thus, the difference in opinion perhaps has to do with what the teacher imagined would have been the form used in the original statement in direct speech.

Teachers reacted differently to cases where some students had changed the verb that was supposed to be used to a verb that is semantically similar. Some students offered an alternative to the verb *murder* in slot 6 (**I ___ (not/murder) him, either**), providing both a milder form *didn't kill* and a more drastic form *didn't slaughter*. Two of the Americans and two of the Finns found the form with *kill* acceptable, but none of the British did. Only two Finns accepted the form with *slaughter*. Furthermore, only two of the Finns accepted using *had become murdered* in slot 4 (**It was soon clear that he ___ (murder)**) and three found it questionable; none of the Americans or British teachers found this form appropriate at all. In slot 42 (**It ___ (normally/take) quarter of an hour to walk to the village**), some students wanted to add the word *a* after the verb to indicate that it seemed to be missing from the next phrase (*quarter of an hour*) in the text, thus providing either *normally takes a* or *takes normally a*. Three teachers found this inappropriate, while two teachers insisted that '*a*' must be added, thus finding *normally takes* without the '*a*' inappropriate.

Despite the fact that some variety-specific tendencies can be found, they do not appear particularly systematic. At times, the Finns and the Americans behave in a similar manner, while at other times, the Finns and the British seem more similar to one another. Much of the variation is slot-specific and difficult to generalise to any systematic phenomenon. Without more evidence, it remains impossible to draw conclusions about the effect of the variety on the assessment. Thus, at least in this study, no systematic difference can be found between the ratings of native and non-native teachers of English, and there is no reason to categorically question the skills of non-native teachers of English, as was done in some studies in Section 2.2. Despite this, some tendencies exist: some Finns were likely to accept forms that native teachers did not accept, but none of these instances included all Finns and no native teachers.

9.2.2 Individual variation

A few of the verb forms caused great individual variation in ways that are difficult to categorise based on any background variable. This section discusses some of the cases where there is considerable disagreement. As explained earlier, in many of the instances where the teachers were not unanimous, the students had suggested several forms, and in many of these cases, the

slot leaves some room for ambiguity. An example is the forms *can't have walked* / *cannot have walked* in slot 106 (**Lucy ___ (not/walk) to the village and back, if ...**): six teachers found them inappropriate, seven the best form. Similarly, in slot 85 (**I ___ (have) a little cottage in the village**), teachers seemed undecided with whether the present or the past tense was better: two found both *had* and *have* the best, three found *have* the best and *had* inappropriate, six found *had* the best and *have* inappropriate and two found *had* the best and *have* acceptable. Some also accepted *'ve got* and *used to have*, others found them inappropriate.

There were a few slots with several competing forms provided by the students that a number of teachers marked as the 'best' or 'acceptable'. These included slots 8, 99, 103 and 104, all of which were discussed above in Section 9.2.1. Thus, disagreement here arose from the fact that some teachers limited their choice of best forms to very few forms, while others approved of several forms as the best. At times, there was a full conflict in the best and inappropriate forms. Furthermore, a variety of tenses and aspects can be considered suitable for example in slots 39 and 40 (**I ___ (never/like) Mum or Dad to be around when I ___ (talk) to him**), which called for great variation in teacher responses. In slot 39, teachers accepted either present, past or present perfect forms, with the British favouring the present, Finns the past and Americans the present perfect. However, individual differences persisted here as well. In slot 40, the suggested forms included the present simple, present progressive, past simple and past progressive, with individual levels of support for the various forms. Slots 39 and 40 are discussed extensively in Sections 8.2 and 8.5.

Teachers disagreed on whether changing the requested verb was allowed. For example, in slot 28 (**My sister ___ (phone) and ...**), students were supposed to use the word *phone*, but a few students decided to use *call* instead. In total, 310 students used some form of *phone*, seven used some form of *call* and two used both (i.e. *phoned/called* or *phoned/call*). Here, teachers disagreed on whether you were allowed to change the verb while keeping the right tense and on whether you were allowed to add an object to the verb (i.e. *phoned me*, *called me*). Thus, three teachers felt both *called* and *phoned* were the best alternatives, while four thought that although *phoned* was the best, *called* was acceptable, and six teachers believed it was inappropriate to use *called* here when the required verb was *phone*. Furthermore, in the case of adding the object, eight teachers thought *phoned* and *phoned me* were equally satisfactory choices, four considered *phoned me* acceptable and one felt it was inappropriate. The form *called me* was rated as the best alternative by two teachers, four considered it acceptable, two questionable and five inappropriate.

Another intriguing case is seen in slot 87 (**... when I started ___ (earn) a lot of money**). Here, a form of the word *earn* is to be used after *started*. According to *Longman Grammar of Spoken and Written English*[111], "the meaning difference between *to*-clauses and *ing*-clauses controlled by aspectual verbs (e.g. *start*, *begin*) is even more subtle [than with some other types of verbs], and in some cases the two seem to be virtually interchangeable" (Biber et al. 1999, 759). However, sometimes there is a difference: when *start* is followed by an infinitive, this "can indicate an intention to begin an action", and when it is followed by an *–ing* clause, this "generally indicates that the event in question has truly begun to happen" (Biber et al. 1999, 759). Textbooks often simply claim that you can use either form: for example, *English Grammar in Use* states that *start* is one of the words that "can be followed by **–ing** or **to** ...", but that "normally we do not use **–ing** after **–ing**", i.e. the *–ing* is rare if the verb *start* is in a progressive form (Murphy 2004, 112). In this study, however, all teachers rated *to earn* as the best alternative after *started*, but only eight considered *earning* the best alternative[112]. Four teachers (two Finns, one Brit, one American) considered *earning* acceptable and, indeed, one teacher (a Finn) thought *earning* was inappropriate. Nevertheless, the context clearly implies that the earning had truly begun, it was not just an intention. Thus, some teachers seem to have acted against the standard verb patterning in English. The students favoured the infinitive: while 96 students provided *earning*, twice as many, 200 students, provided *to earn*.

However, recent research argues that contemporary language use is facing a change regarding the distribution of the infinitive and gerund. Leech et al. (2009, 118-143) discovered that since the 1960s, native speakers have begun to favour gerunds over infinitives and finite clauses with particular verb types, including after the verb *start*. The rise of the gerund is fairly recent, beginning from the 17th century, while infinitival structures have gradually become more widespread than finite ones over a longer period of time. Some of the change is parallel in British and American English, while in some other instances, there are divergent developments (Leech et al. 2009, 186-187).

The verb *start*, which in this context[113] is a catenative verb "used to indicate the beginning, continuation or end of an activity or state" (Leech et al. 2009, 195), allows two types of complementation, the gerund as well as the infinitive. The distribution of these forms in native-speaker usage has long been fairly even, but with the long-term general trend favouring gerunds,

[111] This book uses extensive corpora to describe how the English language is used.
[112] Remember that they could indicate several 'best' forms.
[113] For a discussion of the narrowly lexical and the more abstract meanings of *start*, see Leech et al. (2009, 195-199).

their use has increased. The increase of the gerund is more pronounced with *start* than with the more formal *begin*, and particularly so in spoken language and in American English. Leech et al. (2009, 120-121) maintain that the change has been rapid and the gerund has become grammaticalised quickly, with the process being speech-driven. In statistical terms, the change of forms after *start* is from a ratio of 13 infinitives to 7 gerunds in the 1930s to 49 infinitives to 59 gerunds in the 1990s in British English, and from 47:49 in the 1960s to 59:110 in the 1990s in American English[114] (Leech et al. 2009, 301). However, since the students and the teachers in my study favoured the infinitive with *start*, more research would be needed to examine the current distribution of the use of gerunds and infinitives after *start* and potential changes in the way *start* is used. It would also be useful to research whether teaching materials, for example, have a bias towards providing examples with the infinitive. Another possibility is that the teachers and the students in this study simply are more conservative in their choice of the form after *start*.

Although most students mastered irregular past tense forms, some students provided the regular forms for such verbs. In slot 37 (**... so I ___ (creep) out by the back door**), several students (171) suggested *creeped*. One of the teachers (Teacher 9, who is British) found this the best form, while one found it questionable and the others inappropriate. In contrast, 117 students provided *crept*. *Longman Grammar of Spoken and Written English* (Biber et al. 1999, 394-398) does not list the word *creep* in its list of irregular verbs, but it is mentioned in *English Grammar in Use* with *crept* as the past simple and the past participle, which is also the opinion of, for example, *Macmillan English Dictionary for Advanced Learners* (2002).

Spelling becomes an issue in slot 78 (**I ___ (realise) it was ...**), where the word *realise* was spelled with an *s*, not *z*, in the test. Two teachers did not approve of changing the spelling to *realize*, while eleven accepted the spelling with *z*. Students were not particularly eager to change the spelling, but there were 30 students who spelled their response with a *z*.

There was lack of consensus with the form *spilt* as well. Some students offered it as their response to slot 34 (**He shouted at me because I ___ (spill) a few drops of tea on his desk while ...**), but eight of the teachers found it inappropriate and only five the best form. This form is not listed in the list of irregular verbs in *Longman Grammar of Spoken and Written English* (Biber et al. 1999, 394-398) either, but *English Grammar in Use* (Murphy 2004, 292) argues

[114] No data are available about spoken American English from the 1930s. In the 1960s, the British ratio was 36:52.

that the verb *spill* can be either regular or irregular and that the irregular past form is *spilt*, and the book key expects the form *spilt*. According to Murphy (2004, 292), the irregular form is more common in British English, while the opposite is true in American English; Crystal (2003, 204) thinks *spilt* is rare in American English but that in British English, there is "a great deal of usage variation". Finegan (2004, 367) lists the form *spilt* as a British spelling correspondence of the American *spilled*. *Macmillan English Dictionary for Advanced Learners* (2002) offers both forms. However, four out of six Finns, one out of three Brits and three out of four Americans found the spelling *spilt* inappropriate. Here, it is impossible to know whether the teachers were unaware of such an irregular form or whether they found it inappropriate despite knowing of its existence. It was not common amongst students, either: 260 students provided *spilled*, 30 students *had spilled* and one student *had been spilled*, while only 7 students provided *spilt* and one student suggested *spillt*. This slot is also discussed in Section 8.1.1.

There were two items in the text that could be considered fixed phrases, requiring a particular grammatical form (see also Section 8.6). Before slot 79 (**... it was no use ___ (argue) with him so ...**), there is the expression *it was no use*, after which a form of the word *argue* is to be used. *Macmillan English Dictionary for Advanced Learners* (2002) recommends using the gerund after *it's no use*, and *English Grammar in Use* lists the expression under the ones to be used with *–ing* (Murphy 2004, 126). However, two (Finnish) teachers felt it was inappropriate to provide *arguing* (one of these teachers thought *argueing* was the best form, the other believed *to argue* to be the best), and five teachers thought both *arguing* and *to argue* were the best alternatives. Two teachers considered *to argue* acceptable, one questionable and four inappropriate. It is possible that a change is taking place in the usage of this expression as well, or perhaps lack of exposure to the structure makes the teachers uncertain about the form. Students' understanding of the expected form varied: while 101 students provided the suggested alternative, *arguing*, 158 students used the infinitive, *to argue*.

In slot 110 (**I think it's time we ___ (make) an arrest**), the sentence begins with *I think it's time we*, after which a form of the word *make* is to be used. Ten teachers argued you should use *made*, while three teachers thought such usage was inappropriate and suggested using *make*. Four teachers thought either form could be labelled as the best. Two Finns also accepted the use of *will*, as in *will make* or *'ll make*, and one Finn accepted *should make*. None of the British or American teachers accepted such forms, nor did two of the Finns. *English Grammar in Use*, in its discussion of the expression *it's time*, claims that "[w]e often use this structure to criticise or to complain", and that when using the structure "**It's time you did** something", we mean

that "you should have already done it or started it" (Murphy 2004, 70, emphasis original). It remains unclear whether the teachers who did not accept *made* were unaware of the rule or whether they did not think the expression should be used in such a way. Given, however, that only some of the Finns did not accept *made*, perhaps these teachers were not familiar with the expression. Nor were the students: a minority of the students, 49, provided *made*, while the majority, 200 students, provided *make*. Together, *will make* and *'ll make* were given by 36 students; thus, they were rarer than *made*.

Overall, the teachers in this study were not unanimous in how to react to spelling errors, unusual tense, changing the verb, using the simple instead of the progressive, using contracted forms, unusual word order, adding words (e.g. prepositions), leaving out words (e.g. adverbs) and adding modal auxiliaries. Some teachers seemed to have developed a system to rely on in such cases, while others seemed to react to the items case by case, sometimes with contradictory outcomes. Obviously, it is also possible that they either created a system at a later stage during the process of rating the forms, or they may have changed their mind in the process. Some disagreement among the teachers may also result from teachers not being familiar with particular expressions or structures.

9.3 Teacher assessment and self-reported approach to grammar

Table 7 in Section 6.3 presented information about the teachers' background. One of the elements that has not so far been reported on is the question that asked the teachers about their approach to grammar and marking the forms in the study. In this section, I reflect on the correspondence between what the teachers said in their self-reports and what their marking pattern is like, keeping in mind that what teachers report they do "should not be presented as evidence of what they *actually* do" (Borg 2015, 495, emphasis original). Here, I follow Borg's advice and combine the questionnaire responses with an analysis of how the teachers rated the students' responses. This is a partial attempt to answer the question why there is variation in the teachers' ratings and how that is manifested.

Table 38 lists the teachers' responses to the question inquiring about their attitude to grammar and to the (quasi-)question asking if they wish to give any other information that might be

important or useful for the study. The responses are quoted verbatim, including spelling errors or any other unconventional uses of English. The responses were written by hand, and thus, if there was any emphasis used, it was in underlining. This has been retained in the table[115]. As explained in Section 6.3, Teachers 1-6 were Finnish, Teachers 7-9 British and Teachers 10-13 American.

As can be seen in Table 38, some teachers express a positive attitude to grammar: this is at least Teachers 1, 3, 5, 8, 13 use terms such as 'enjoy' and 'interesting'. In contrast, Teachers 4 and and 12 seem to hold negative attitudes, using terms such as 'not interested' and 'dull'; Teachers 6 and 10 are slightly more difficult to interpret but, on the whole, they seem to be more negative than positive. Teachers 2, 7, 9 and 11 do not use words that would imply what their attitude to grammar is, and that was not directly asked, either. Thus, there seem to be slightly more teachers in this sample who take an interest in grammar than in some other studies conducted in Western countries (see Section 3.1), although it may be that some teachers with even more negative attitudes did not even consent to take part in the study.

Table 38. Teacher attitudes to grammar and approaches to evaluating the student responses, given verbatim (spelling errors and emphases original)

Teacher	**How would you describe your attitudes towards learning, teaching and applying English grammar?**	**Any other information that might be important or useful**
1	Positive; descriptive rather than prescriptive	-
2	It's a descriptive tool for communication; All depends on the context and what you want to say.	Always difficult to react to answers in which the student has used xtra words or done something contrary to the task itself -> answer might be correct but it doesn't correspond to the task; dilemma what to do!
3	I enjoy both teaching & learning more about grammar & applying it in exercises. Mistakes bother me (a bit) in writing, but in speaking I feel communication is more important than correctness.	-
4	Not particularly interested but definitely appreciating the importance.	I felt like explaining why I was marking most alternatives with 4: It is obvious that many decisions have to do with issues like written vs. spoken language, "technical" issues like when they repeated a word (not part of the verb) that was actually given in the text. (Was this the other way round in fact?) I am somehow not happy with all my 4's but the other alternatives weren't suitable either (3, when specified as one not being sure if it can be used).

[115] Some teachers responded by using uppercase letters only, but this effect has not been retained.

Teacher	How would you describe your attitudes towards learning, teaching and applying English grammar?	Any other information that might be important or useful
5	Very interested and keen	-
6	It is something that has to be dealt with, it is part of the whole. When grammar is understood as awareness of language rather than showing what is right or wrong, I find grammar rather interesting. I find this "right-wrong" division rather problematic as often right and wrong form a continuum.	I found your task very difficult and had to adopt a stricter approach than I would under normal circumstances where I try to see why something is wrong (e.g. here some students had clearly forgotten to write down the main verb when filling in the auxiliary because it stood there in brackets). On the other hand, I tended to accept forms that seem to be rather colloquial because the texts were from a spoken context although they were in print. I have (I think) accepted the use of the present tense although the people in the task talk about the past. I've done this because real people often resort to the present when talking about the past. However, accepting the present tense is perhaps not always the right choice because what is right really depends what a particular student wrote in the gaps before and after. I also feel I treated certain choices unfairly, because the student might have used a correct form but not the right (i.e. given) verb (cf. phone, ring) – but I had no other way of being able to finish this!
7	Those of a teacher - pedantic	Being a teacher, a lot of one's evaluation of the test responses is in terms of what she wants a test to do, and knowing what the textbook writer is "getting at" in devising these particular items. Hence one may be stricter about some things, but less strict about others (e.g. allowing both contracted and uncontracted forms, or allowing respondents to include the following word – see e.g.36) than lay-persons.
8	I enjoyed English grammar lessons at school. Teaching + paying special attention to grammar in English depends v. much on the the context. On the whole I prefer to put the emphasis on communication but I do point out the usefulness of good grammar + the occasional need for accuracy.	My attitude to grammar mistakes also depends on the context. It is always important to make criteria (for assessment, for eg.) transparent and clear in advance. In the Exercise done here, I took a fairly strict approach but on another occassion (say, working through an exercise like this with a group) I would explain other possibilities (and the issue of spelling).
9	It's difficult to say – do you mean how important it is for students to get it right? Again – it depends on the situation: where lack of grammatical accuracy impedes communication, it is clearly a problem.	-
10	I'm not a big fan of English grammar. I am frustrated by the often lack of logic. Nevertheless, I see importance in it, and enjoy teaching it when there is logic. Still, I put more importance on meaning instead of form.	I sometimes had trouble marking my answers – I tried to really strict – e.g. putting "4" when the spelling wasn't correct or when a new word was introduced that was not in the originally text. Occassionally I may have not been completely consistent though.

Teacher	How would you describe your attitudes towards learning, teaching and applying English grammar?	Any other information that might be important or useful
11	The challenge is between spoken & correct English and making that distinction for students. Grammar must be learned from a book, not listening as people make to many mistakes in speech.	Reading aloud can help students practise the correct forms.
12	I'm not strict = perscriptive – vs – conscriptive grammar – who cares? Grammar can be useful but strict grammarians are somewhat dull & outdated.	-
13	I actually think grammar is interesting! While aware of the fluidity of the structures over space and time, I'm also fairly conservative and try to preserve stability, and so tending to think in terms of "correct" and "incorrect".	-

In the form where he marked the student responses, Teacher 1 commented on a problem that is, in his opinion, a matter of principle: if a student did not use the verb that was supposed to be used in the slot, should the teacher accept all forms that are lexically appropriate in the context? When analysing the responses, we see that he decided not to accept them. He also underlined the forms he would use himself but occasionally also marked other forms as the best or acceptable and crossed out the items after he had rated them. Teacher 1 was somewhat critical of using contractions and, at times, found them inferior to uncontracted forms, marking them with 2 (acceptable) or 4 (inappropriate), although at other times he found them equally acceptable to the uncontracted forms, particularly in the first few slots. He rarely used the 'questionable' category and tended to mark a lot of answers with the label 'inappropriate'.

Teacher 2 had a similar approach to the forms, although he did not have a tendency to favour uncontracted forms. Incidentally, in a question written into his answer sheet, he wondered if we should take into consideration how a response to a prior slot affects the response to a later slot, for example, from a present tense to the present perfect and from the past tense to the past perfect. In his grammar approach, he mentioned the difficulty of knowing what to do when students had added extra words or responded in a manner that the task did not expect them to do. He also wondered what to do if the answer was correct as such, but not the one expected for the slot. In the responses, he seemed to treat these case by case, accepting for example *called* as well as *phoned* but not *rang* in 28 (**My sister ___ (phone) and we ...**). Teacher 3, in contrast, used the questionable category much more extensively than Teachers 1 and 2, and she used category 2 (acceptable) extensively as well. Thus, she seemed to build a continuum between 'right' and 'wrong' to a greater extent than the first two teachers.

In the verb list, Teacher 4 explained that she made a policy of always marking the forms with an extra preposition as unacceptable, but she wrote that she wondered about this, since often the verb form itself was correct; the students just did not notice the preposition in the text. In the grammar approach, she explained that she decided to use the category 4 (inappropriate) extensively, because she felt that category 3 (questionable), was not suitable, perhaps because the questionnaire explained that this category could be used if the teacher was unsure if the form could be used or not. It is possible, of course, that if the category had been worded differently, she might have used it more; nevertheless, she seemed to feel generally uncomfortable with the task and was not happy with the strategy of "overusing" the category 'inappropriate'. Teacher 4 did not use many 2s (acceptable) or 3s (questionable) in the marking. Teacher 5 used the 'questionable' category but was perhaps a little cautious with category 2 (acceptable).

In her additional comments, Teacher 6 explained that the form *the shouting* in slot 19 is not actually a verb and found it inappropriate. Generally, she followed a principle of dismissing forms where students had not used the verb that was intended in the particular slot. She used the full range of the rating scale, but perhaps the most so with marking forms more readily with 1, the 'best'. She said she adopted a stricter approach than she would otherwise apply. However, she accepted colloquial forms to match the spoken context and the present tense for talking about the past. She notes that what is 'right' depends on the student's choices before and after a particular slot and felt that she did injustice to some students for punishing them for not using the given verb, although the form was correct. As Table 38 tells us, she was uncomfortable with the task and having to mark the forms, and perhaps the use of the full range of criteria was her way of trying to show the existence of a continuum instead of a dichotomy in the forms. Perhaps her comments in the grammar attitudes section indicate that she felt a sense of dissatisfaction with the task because she was not able to explain her rationale or because she found having to use the rating scale limiting and discouraging. Like Teacher 4, she seemed to have completed the task reluctantly.

Teacher 7 decided to write down in the verb list what he thought the answer should be, but he occasionally also marked other choices than his own with the 'best' label. He commented that he was occasionally perhaps stricter and occasionally less strict with some instances in the task than what the writer of the textbook had expected, or what lay-people might say in these instances. He did not tend to use the categories 2 (acceptable) and 3 (questionable). He also soon started to mark items that he considered inappropriate groupwise instead of marking each

form separately; Teachers 2 and 12 did this as well, but only Teacher 7 commented on having decided to follow this practice in the verb list. Perhaps these teachers' decision to "mass-mark" forms 'inappropriate' discouraged them from considering further options extensively.

Teacher 8 decided to take a fairly strict approach, hoping to establish transparent criteria, but she also wrote that in teaching contexts, she might accept other possibilities and explain these to the students. She barely used category 3 (questionable) and, at times, she also used category 2 (acceptable) sparingly. However, in some instances, she found several 'best' alternatives. For example, in slot 99 (**It ___ (not/be) his wife**), she marked 12 forms as the 'best'.

Teacher 9 followed her own principle of "it depends on the situation" and used the scale fully, although she did not have many 3s (questionable). She marked several forms with 2 (acceptable) and at times also with 1, the 'best'. She seemed fairly similar to Teacher 3 in her approach.

Teacher 10 said he wanted to be strict in marking the forms, for example misspelled words and using a different verb than the one given, and indeed, he used the inappropriate category quite extensively. Despite this, he also acknowledged he might not have been very consistent with this after all and, at times, he indicated several 'best' or 'acceptable' responses. He occasionally commented on some forms with "maybe in spoken" or "I think only in spoken". Actually, Teacher 11 also occasionally labelled forms with "sp", which perhaps refers to spoken form, although this remains unclear; she tended to mark these 'sp' forms with 2 (acceptable). She also occasionally wrote "did not follow instruction" in the list, for example when students had added words to the verb forms, as for example the word *with* in *not in love with* in slot 5 (I ___ **(not/love my husband, he was a cold and selfish man**). Teacher 11, who believed that students should learn their grammar from written, not spoken English, occasionally used both 3s (questionable) and 2s (acceptable), but not extensively. Perhaps this reflected her thinking that standard written forms constitute the rules in grammar. However, the task represented spoken English in written form, which may be a complicated genre to rate.

Teacher 12 used the full scale very thoroughly, similar to Teacher 6. However, while Teacher 6 seemed to want to display the fact that there is a continuum, not a dichotomy in grammatical choices, Teacher 12 seemed to find a focus on grammar old-fashioned and dull. Thus, his lenient approach to students' forms perhaps reflected his lack of strictness: several forms were given as both the best, acceptable, questionable and inappropriate. However, he did not seem to think

that “anything goes”; rather, he seemed to have considered the forms carefully. He tended to use category 3 (questionable) quite extensively, while he had few 2s (acceptable) towards the end of the list. Despite saying that he was not strict, he seemed at least somewhat stricter than Teacher 6. Teacher 13, on the other hand, thought he was fairly conservative and tended to think along the dichotomy of “the best” and “inappropriate”. He did not really use the ‘questionable’ category, but occasionally resorted to category 2 (acceptable).

Based on the above, the teachers could perhaps be grouped based on their approach to rating the acceptability of the students’ forms. Teachers 1 and 2 (Finnish), 7 (British) as well as 10, 11 and 13 (American) seemed to seek a simple solution, tending to either accept or disapprove. They could perhaps even be called strict in their decision-making within the context of these 13 teachers.

Teachers 3, 4 and 5 (Finnish) as well as 8 and 9 (British) could perhaps form a group that falls somewhere between strict and lenient; they could playfully be called the “it depends” group. However, there is some variation within the group. Teachers 3, 5 and 8 were a little more open for “inbetween” forms, but they were not as open to novel approaches as some other teachers. Teachers 3 and 5 tended to favour the ‘questionable’ category, while Teacher 8 preferred to extend her assessment toward labelling more ‘best’ forms. Teacher 4 could be characterised as somewhat hesitant. Teacher 9 seemed to be willing to accept more variation by rating the forms with ‘acceptable’.

Teachers 6 (Finnish) and 12 (American) seemed to be the most willing to allow for flexibility and freedom in the use of grammar, although they did so for slightly different reasons. They made full use of the entire rating scale and readily accepted alternative forms. Teacher 12 stressed the fact that he was not strict and did not really care about grammar, while Teacher 6 highlighted the significance of the context and the versatility that exists in the “real world”. She said she would rather focus on language awareness in her discussion of grammar.

The approach the teachers adopted influenced how likely they were to search for alternatives beyond the categories ‘the best’ and ‘inappropriate’. The choice between leniency and strictness seems to have been a quality that the teachers had adopted either consciously or unconsciously, and it seems to have reflected their approach to grammar as a whole. This did not seem to correlate with age, variety or level of education, but there were more strict men than women,

and the most experienced teachers can all be categorised into the "it depends" category, falling between the strict and lenient approach.

Overall, variation in the teachers' rating mainly focused on the extent to which they decided to make use of the categories 'acceptable' and 'questionable', and to the extent to which they were willing to accept alternative interpretations of the context. Some of the variation may be attributable to ongoing changes in contemporary English, while some is also explained through varying decisions on how to react to changes initiated by students in the form of adding words, omitting them and changing the verb. The variation seems to follow individual decisions rather than to be bound to either variety, gender, age, experience or other background factors, while it is noteworthy that the most lenient Finn was the one who had spent the longest time in an English-speaking country. Nevertheless, the strict teachers include both Finnish, British and American teachers.

9.4 Summary

The teachers also exhibited a range of variation in their assessment of the students' responses. They used a four-part scale to rate the responses, including the best, acceptable, questionable and inappropriate category. There were 13 teachers, who all taught English at the University of Helsinki. Six of them were Finns, three British and four American. They were from 31 to 63 years of age, and the majority had several years of teaching experience at the university level. Seven of them had a Master's degree, while six had either a doctoral or a Licentiate's degree. They each rated 1,522 verb forms and provided a total of 19,786 ratings, including 2,258 ratings as the best, 867 ratings as acceptable, 439 ratings as questionable and 16,151 ratings as inappropriate. Thus, 84% of the forms were found inappropriate or questionable and therefore considered errors.

The teachers fully agreed on 64 forms as the best ones for their context and unanimously dismissed 825 forms as inappropriate. Thus, there was variation in the assessment of 42% of the verb forms. Quite often slots with considerable student variation also proved to reveal teacher variation. There were 203 cases where only one teacher disagreed with the other 12 teachers and 174 cases where two teachers disagreed with the remaining 11 teachers.

The teachers can be grouped according to two variables: how often they were the only one to disagree with the remaining 12 teachers and how likely they were to be strict or lenient, which showed particularly in the number of acceptable and questionable responses and occasionally in the acceptance of several 'best' options. Some teachers allowed for more variation in students' responses, while others were more likely to adhere to more conventional assessment. The teachers' level of strictness was also compared against their self-reported views on grammar.

While most of the variation was individual, there are some tendencies that can be traced to whether the teachers were Finnish, British or American, but this was never fully systematic. Some Finns exhibited an increased acceptance of the emphatic *do* in the past tense, i.e. *did*. In some slots, variability was increased by the fact that the context readily allowed for various interpretations. Other reasons included individual decisions on how to treat, for example, missing words or spelling errors and when to accept unconventional tense.

10 Discussion

As we have seen, variation arises from multiple sources. Studying variation in the use of English among learners of different first languages is important to understand to what extent learners' L1 influences the variation. This study has researched variation regarding Finnish learners and English verb forms. Investigating where the limits of acceptable variation lie and how competent university students are in using English verb forms is also important, as this can help us understand where the focus of instruction should be. Because this study has explored variation from various perspectives, the discussion has four parts. The first two sections discuss variation within the student population, beginning with variation in the provision of different forms (Section 10.1) and continuing to variation in students' skills and the background factors that best account for success in the test (Section 10.2). Section 10.3 focuses on variation in the teachers' responses, and Section 10.4 explores implications arising from the study.

10.1 Variation in verb forms

This section answers the first part of the first research question, "*What is the extent of variation in Finnish university students' use of English verbs in a fill-in-the-gap test, and what accounts for this variation?*" Obviously, no simple answer exists. The extent of variation depends on the particular slot; however, arising tendencies are discussed below.

The test proved successful in providing room for variation, yet limiting it to specific contexts. It formed a contextualised entity in the form of a detective story, with 107 slots for students to fill in with verb forms. In the majority of cases, the expected answer was the past tense, with the past simple expected in 50% of the slots, and the past progressive in 15%. These were followed in frequency by the past perfect (10%), verbs with a modal auxiliary (6%) and the present simple, *to* + infinitive and the gerund (5% each). Thus, past forms greatly dominated the test.

Overall, Finnish students are most comfortable using the simple past tense. The affirmative past simple yielded the highest scores, even reaching 98% in several slots. Accuracy was greater with the affirmative, because spelling errors increased with the negative forms. Additionally, a

handful of students provided the past negative with a so-called double past, with both the auxiliary and the main verb taking a past form (e.g. *didn't went*). Thus, the simpler the form, the more accurately it was used. However, some discrepancy existed in the treatment of contracted vs. uncontracted forms because some teachers did not always approve of the use of contracted forms.

While Finnish students score well with the past simple, more variation occurs with most other verb forms. However, Lehtonen and Sajavaara (1983, 111) suggest that a situation where students can monitor their responses can cause more variability than other situations. The testing situation possibly made students seek alternative forms to a greater extent than, for example, a discussion context or a free-writing exercise would have done; however, the situation possibly made the students choose 'safe' options. Alternatively, given the prevalence of the past simple as the expected form, perhaps students routinely continued to provide the past simple without contemplating their responses to the extent they should have.

The success rate with forms other than the past simple remained low. The average score for providing the past progressive in slots where it was expected was around 50%, while the success score with the past perfect was only 43%. Several students seemed unfamiliar with the requirements of reported speech, which affected some of these forms. As explained in Section 4.4.2, some error types, such as tense, verb complementation and finite vs. non-finite verb forms, showed no statistically significant improvement across CEFR levels in a study by Thewissen (2013). In my study, a similar finding is made: despite years of study, many students struggle with the progressive and the past perfect. Thewissen (2013, 93) argues that tenses remain a difficult part of English grammar, perhaps because students need to become proficient in both tense and aspect, and because tense choices often have implications for larger elements than single clauses. However, if Finnish university students are, as assumed, at level B2 on the CEFR scale, they should be able to distinguish between contexts requiring the simple and the progressive as well as the past and the past perfect, and their struggles imply a need for more intensive instruction in such forms. Although the CEFR avoids discussing grammatical forms in its descriptors, the B2 level already presupposes a strong command of the target language. For example, B2 level writing must be clear and detailed, and students should be able to write essays and reports (CEFR 2001).

Although a number of recent studies have shown that the use of the progressive is spreading beyond the traditional uses (see Section 4.2), such a phenomenon is absent from my study. This

absence may relate to Finns' struggles with even the conventional uses of the progressive: even English Majors had some difficulty providing the progressive in contexts where it was expected. Students possibly obtained different scores from using the progressive due to the prototypicality of the verbs (e.g. L. Collins 2005) or the extent of exposure to such forms. Furthermore, no slot in the test particularly called for the use of the so-called new progressive.

Interestingly, few students used any modal verbs, even though modal verbs were the expected form in some slots. Students may not have considered modal verbs relevant to the task, or some may have avoided modal verbs on purpose. Additional complications in the use of verb forms arose from verb complementation patterns and fixed phrases; in other words, increasing structural complexity of a verb phrase increased variation. Furthermore, some students had difficulty with word order in the verb phrase. Of the two slots that required the passive, students scored better regarding the past perfect slot than with the past continuous, which also required attention to word order in the verb phrase. As discussed in Section 4.3, Hinkel (2002; 2004; 2011) has also found that non-native university students would benefit from further instruction in the passive. Some variation was also caused by adding or omitting words (cf. Section 3.2.1), changing the verb or overusing the emphatic *do*, which was never the expected form in the test.

Some systematic variation (see Regan 2013 in Chapter 2) was found in the choice of using contracted or uncontracted negative forms. Nonetheless, contrary to Regan's (2013) claims, not all variation phenomena were systematic: some students fluctuated between correct and incorrect spelling, while other students used both contracted and uncontracted forms without any particular pattern. Spelling errors already appeared to be fossilised in some students. However, Regan (2013) focused on variation in oral use, while this study addresses written grammar, which can explain the difference between the results.

This study provides some interesting results regarding the frequency of using contracted forms. Although students often used contracted forms in the negative, their use in the affirmative was rare. Interestingly, English Majors were the least likely to provide contracted forms. This discrepancy may result from greater exposure to contracted forms in the negative, or it may reflect some uncertainty in how to contract affirmative forms. One further possibility is that students had been advised against using contracted forms or that they felt that, in a test setting, they should provide the full forms. However, some of the negative sentences had a prevalence of contracted forms, up to 70%, and it would be quite surprising if such advice only affected affirmative clauses. Thus, the hypothesis regarding greater exposure seems more credible.

While Hawkins (2007) and Hawkins and Casillas (2008) suggest that students may not perceive the marking of grammar in words that are not stressed or in forms that have been reduced if their L1 does not prime them to do so, no such tendency appeared in my study. Actually, while the Finnish language does not have word-final consonant clusters, few students in my study failed to provide the regular past ending (cf. Hawkins 2007): on the contrary, the past simple was the form they were the most confident with. However, some students dropped other words, such as prepositions and adverbs, which may indicate that they felt such words were redundant (see Section 2.3.2). Similarly, students' underuse (cf. Ortega 2009, 41-42) of the progressive and the past perfect may result either from avoidance, if students were not confident with such forms, or from transfer from Finnish. This latter explanation could account for the difficulty with the progressive but not with the past perfect, as the latter is marked in Finnish as well. Finnish students were also more competent with regular than irregular verbs, which contradicts the order of acquisition assumed true for all learners (see Section 4.4.1). Nevertheless, as the students were already at an advanced level, it is possible that they had been more likely to use irregular verbs successfully earlier in their acquisition process. The Finnish language, however, does not have irregular verbs, a fact which has traditionally been considered as resulting in Finns being better at using regular verbs and easily overgeneralising the *–ed* ending. The assumed order of acquisition should perhaps be revisited to ensure that there is no bias from particular first languages.

In summary, although the students were mainly successful with providing the past simple, they need to focus more on progressive forms and the past perfect, as only half of the students were proficient in these forms. These forms are also readily teachable, while it is more difficult to teach the patterning of individual fixed phrases because they are not as easy to generalise. However, further attention to verb complementation patterns would also be useful. Overall, students need more awareness of the wealth of forms that are available and what contexts require their use, so that the students would venture to use forms beyond the past simple.

10.2 Variation in students' skills

This section answers the second part of the first research question, "*What is the extent of variation in Finnish university students' use of English verbs in a fill-in-the-gap test, and what accounts for this variation?*", and focuses on the background factors that affect students' success in the study.

The students in this study came from various backgrounds, but they all had either Finnish or Swedish as their L1, had been educated mainly in Finland, had started their English studies in primary school as either their L2 or L3 and had completed the matriculation examination before entering university. Their average age was 22 for women and 23 for men, and they majored in a variety of disciplines as students at the University of Helsinki, Helsinki University of Technology or the Finnish Academy of Fine Arts.

In the test, the criteria for errors were based on teacher evaluation and the book key, and different levels of strictness were created. In the most lenient approach, any form that at least one teacher considered either the best or acceptable became a correct answer, and errors were only the forms that all teachers found either questionable or inappropriate. The moderate approach required the acceptance of at least four teachers and the conservative approach at least ten teachers. The strict approach required either the acceptance of all teachers or an answer given in the book key. As demonstrated earlier, the teachers did not always agree with the book key or with each other.

The results provided in Chapters 7 and 8 informed us of two aspects: the students' overall skills and the range of variation in their responses. Although a fairly high number of students were proficient in the use of verb forms when the lenient approach was used, the number of inappropriate forms increased when more teacher consensus was expected. As discussed in Section 4.3, Hinkel (2002; 2004; 2011) finds non-native speakers' skills in English grammar and academic writing lacking. While she can be considered too demanding, this study supports her arguments by considering further attention to grammar necessary. While the test employed in this study is not from an academic setting, the participants need to engage in academic writing in their studies, and their lack of proficiency in verb forms could make this challenging.

As we have seen, students' mean numbers of errors vary depending on which criterion is used; tightening the assessment criteria quickly leads to more errors. The mean number of errors is 8.75 with the most lenient approach and 31.3 with the strict approach. This range of errors means that the percentage of 'wrong' answers varies from 8% to 29% in the entire student population. The move from the lenient to the moderate criteria almost doubles the number of errors (from 8.75 to 16.25), as does the move from the moderate to the conservative (from 16.25 to 30.6), while the difference between the conservative and the strict is only 0.7 errors. However, some students made up to 70 errors, which represents 66% of the slots. Thus, they were more often wrong than right.

Interestingly, no slot existed where all students answered correctly even with the most lenient approach. Then again, as indicated in Table 20, all students automatically "made" three errors with the conservative criteria, simply due to the lack of teacher consensus on the correct form. This lack of consensus resulted from slots that could be interpreted in different ways.

The variety of forms produced by students is very wide overall, although narrowest among students majoring in English. While both proficient and weaker students provided verb forms that were unusual, the weaker students mainly offered unusual spellings or mixing of tenses. In contrast, high-achieving students were more familiar with unconventional, yet acceptable verb forms. Of course, students with good skills may be more confident and more willing to experiment with less frequent forms. Consequently, they could benefit from the possibilities of choice afforded by good skills in grammar (see Larsen-Freeman 2002; Celce-Murcia 2002).

The background variables that best explain success in the test are, unsurprisingly, the student group and previous marks in English. Many of the students with the best school marks were English Majors, and on average, they scored the best. They were followed by the Technology group, while the Visual Arts students had the highest error average per student. However, some individual Grammar group students made the most errors. The students with the least errors had already succeeded at school, with mainly the top two marks both at school and in the matriculation examination. However, only the top marks had a more direct association with good test scores, while marks from 8 or *magna cum laude approbatur* downwards showed more variation. The students with the lowest school or matriculation examination marks did not score the lowest in this test. In particular, students with either mark 8 or *magna cum laude approbatur* scored either very well or poorly: these marks were the most unpredictable regarding test success. This unpredictability is somewhat worrying from the equality perspective, when much emphasis will be placed on matriculation examination marks in future Finnish university admission systems, because such marks are not always a reliable indicator of students' actual skills.

Tremblay (2011) found a significant relationship between fill-in-the-gap test scores and the broad proficiency estimates derived from students' background information. As we can see, a similar phenomenon takes place in my study: the students' marks from school and the matriculation examination strongly correlate with success in the study. This may indicate that such examinations largely test accuracy. Furthermore, students may be accustomed to fill-in-the-gap tests, as they tend to be a popular exercise type at school.

My own pilot study (Pesola 2002) showed that students who mastered several second languages, who had good experiences from school and who reported a positive attitude and motivation to study English had also achieved good marks at school and scored well in the test. In the present study, students with good school marks and knowledge of several languages also scored better than students who were less interested in languages. Although unsurprising, it confirms the strong link between motivation and success. In this study, many more women than men had studied several languages, and many English Majors in particular spoke several languages.

The students who studied a language major other than English were a little better than students with non-language majors, but the difference was not statistically significant. Students who had started English as an L3 were also a little more proficient than students with L2 English, although again, the difference did not reach significance. The students who had English as their L3 may have benefitted from their L2 in the acquisition of English (see Jarvis 2015; Klein 1995; Ortega 2009 and Section 2.3.1). Interestingly, English was the L3 for many English Majors, which may indicate that they were interested not only in English but in languages in general and had perhaps a higher aptitude. People who take an interest in languages are possibly more ambitious in studying languages overall, and they may have more motivation to aim at both accuracy and fluency. Different student groups naturally can have different interests in life and studies. The Visual Arts students, whose average score was the lowest, may not have invested much effort in English grammar at school but instead focused on something else.

Although previous research (see Section 5.4.2) discovered a benefit for L1 Swedish speakers in the acquisition of English, this study found no statistically significant difference between L1 Finnish speakers, L1 Swedish speakers or bilingual students. This may relate to Swedish-speakers not benefitting as much from crosslinguistic transfer in verb forms as in some other aspects of grammar. Additionally, this study focuses on the advanced level, where differences caused by the first language have largely been levelled out (Palmberg 1977). However, comparing L1 Finnish and L1 Swedish students' skills is difficult, because only a limited number of Swedish-speakers took part in the study.

No correlation with test success was found for several factors: for example, older students were as successful as younger students, and students who had taken additional courses in English or spent more than a month in an English-speaking country did not outperform those who had not done either. This study found no difference in scores between men and women, while in some

previous studies, Finnish girls scored better than boys (see Tuokko 2000 and Section 5.4.1). The Technology group scored the second-best and was male-dominated; one possible explanation is that increasing consumption of media and the Internet have bolstered men's accuracy rates. However, the Computer Science group was also male-dominated, but did not score equally well, implying that this theory requires further study.

Students' use of English outside educational settings does not seem to boost their skills in grammar. In this study, active engagement with and productive use of English did not correlate with success in the test, while there was a slight benefit for students with passive (receptive) consumption of English, particularly for students who read books in English. Some previous studies have also reported that reading literature in a foreign language results in greater accuracy (Tuokko 2000). Such 'passive' use of English was mainly reported by men studying either Computer Science or Technology, while many of the active users were female English Majors. An increased visual exposure to English possibly helps students improve their accuracy, particularly when this is initiated by the students' own interest. Time invested in reading may also provide ample opportunities for students' immersion in versatile grammar; people who enjoy reading possibly tend to focus more on the way language is used.

Similar to Alanen (1997), a comparison between students' scores and their background variables suggests that extended visits to English-speaking countries do not affect students' performance in the test. Furthermore, no statistically significant link existed between studying additional courses in English after upper secondary school and the number of errors. Thus, students' scores in this test are not related to the extent of their attempts to improve their English skills beyond or after school. The benefits of frequent encounters with English seem more likely manifested in students' lexical skills, fluency and oral skills than in grammatical accuracy. Obviously, such improvement in students' ability to communicate is important, but it does not assist students in gaining superior skills in accuracy.

10.3 Variation in teachers' evaluation

This section answers the second research question, "*What is the extent of variation in teachers' responses to the variation displayed by Finnish university students in their use of English verbs, and what accounts for this variation?*" Again, the simple answer is that the extent is wide, but

limitations exist regarding the variation allowed by teachers. The reasons for the variability are discussed below.

As we have seen, there were 13 teacher participants, who were asked to rate the students' responses on a scale of 1 to 4, with 1 indicating 'the best' answer, 2 'acceptable', 3 'questionable' and 4 'inappropriate'. The teachers rated 11% of the forms as the best, 4% acceptable, 2% questionable and 82% inappropriate. Thus, about 15% of the forms were found to be 'right' and 85% 'wrong'. Notably, the teachers did not often resort to the middle of the scale: only 6% of the forms were rated as 'acceptable' or 'questionable'. This behaviour contradicts findings in some other studies (see Section 3.3), which argue that teachers tend to apply a continuum rather than a dichotomy in their assessment. Nevertheless, certain teachers were more eager to use the 'continuum' approach and are responsible for a bulk of the 'acceptable' and 'questionable' assessments. The gradient scale, with four points, allowed using a more refined assessment if the teachers wanted to engage in such activity. Thus, some teachers used the full scale, while others were more inclined to use only the extreme ends of the scale. The test context was not rigid (cf. Boas 2011), and some slots allowed various understandings of the expected verb form. As this study investigated variation in teachers' assessment, the participating teachers themselves set the norms for the rating, which were partly dissimilar.

Despite the extent of variation in teachers' responses, significant consensus also exists: of the 1522 forms, the teachers were unanimous in rating 889 forms, which constitute 58% of the forms. In particular, they were unanimous in finding 825 forms (54%) inappropriate and 64 forms (4%) the best. Therefore, disagreement concerned 633 forms. Of these, only one teacher disagreed with the remaining 12 teachers on 203 forms (13%) and only two teachers disagreed with the remaining 11 teachers on 174 forms (11%). Thus, the more extreme variation found in the study is limited to 256 forms, which represent 17% of all the slots. Obviously, these are the most interesting cases for my study, as I focus on variation, but it is important to remember that the teachers fully agreed on more than half of the forms. Thus, the influence of lenience or strictness is limited to a number of forms that are debatable and a number of slots that are ambiguous. The teachers were mostly unanimous about more radical spelling errors and the creation of double past forms: these were typically considered inappropriate.

However, the findings show some variation and disagreement over the use of verb forms. The differences in teachers' opinions do not seem to follow alleged native vs. non-native teacher lines (see Section 3.2.2); instead, they are more individual. Additionally, the teachers in the

study did not always agree with the book key. Of the 107 slots, teachers were entirely unanimous with the key in 60 answers; for three slots, no verb form gained a convincing majority. These ambiguous slots mainly divided opinion due to teachers' differing inclinations to accept or seek alternative interpretations to events. Furthermore, as noted by McNamara (2000, 37-38), teachers, being human, are not always systematic: while some consciously tried to apply systematic criteria, others explored each case separately without relying on any pre-existing criteria. Similarly, some teachers more readily accepted alternative ways to perceive a context; again, this was more systematic for some teachers than others (see also Huhta et al. 2014). Thus, teachers differ in how sensitive they are to alternative interpretations, which may be influenced by how much time they are willing to invest in rating the answers. As considering all potential approaches to a specific slot is both time-consuming and intellectually challenging[116], some teachers possibly decided to cut corners and only acknowledged the most obvious solution.

As we saw in Section 7.3.2, an interesting phenomenon occurs with the conservative criteria (when ten or more teachers must agree on the form being either the best or acceptable): in three slots, no response is the 'correct' form. The context of these slots allows for multiple interpretations of event time, some of which could pass unnoticed within course settings, for example, and only become evident in a research setting. As Lehtonen and Sajavaara (1983, 111) point out, a situation where students are specifically asked to monitor their behaviour and where teachers are given a scale to use for research purposes may cause more variation in the responses than, for example, marking or commenting on students' papers within a course or test setting. However, while students tended to have more variation with the more complex forms, teachers tended to expect command of both the more complex and simpler forms.

Most of the teachers in the study expected standard spelling, with limited allowance for variation. Most forms that were spelled in an unconventional way were rated as 'inappropriate'. This rating may reflect an assumption that students who have reached the university level should already master spelling, particularly with frequent forms such as *didn't*. Nonetheless, some teachers practiced some leniency towards unconventional spelling, but no systematic pattern seems to arise in the type of spelling errors they accepted. This absence causes additional variation in the teachers' responses.

[116] One teacher participant decided to withdraw her participation after having rated the first 40 slots, exclaiming that the task became impossible as she would have had to keep too many options in mind. Her responses are not included in the study.

Some assessments provided by the teachers may have been based on, for example, the expected level of formality for the test. However, this basis is not evident in the responses in a systematic way, although some teachers seem to dislike the use of contractions. Since the test is not clearly either spoken or written language, deciding which conventions to follow may have become difficult. The decision on the acceptability of a form may have been based on the interplay between conformity with conventional spelling, tense and acknowledgement of nearby forms.

Huhta et al. (2014) found that removing a teacher with a radically different scoring pattern from the rest of the teachers did not affect the overall assessment. In contrast, the difference between different levels of strictness resulted in radically different error scores in my study. However, in my study it was not always the same teacher who rated the forms differently from the rest, but different teachers rated different slots and forms in varying ways. Neither of the two most lenient teachers can be held responsible for the variability in the assessment, since all teachers found some forms acceptable when the others did not.

The variation in teachers' assessments in this study conforms with Huhta et al.'s (2014) findings: no particular pattern can be established, but the leniency or strictness teachers adopt is connected to their individual preferences. Hyland and Anan (2006), however, found that Japanese teachers rated errors differently to native English teachers; perhaps this reflects a cultural difference between teachers in Asia and those in Europe (see Section 3.1.2). In my study, although groupings based on nationality can be established for a few specific verb forms, these remain rare and almost always come with exceptions; the vast majority of the ratings reflected individual preferences. However, more men than women were strict in this study, and the most experienced teachers were neither lenient nor strict. Investigating these two factors further would be worthwhile to establish whether this phenomenon is simply a coincidence or whether it reflects more widespread tendencies.

Although the Finnish teachers in this study do not stand out as being either stricter or more lenient at the group level, some verb forms exist where their assessment slightly deviates from both the American and British teachers. This difference may result from Finnish teachers having acquired English as an L2 in an instructional context, while native teachers have become socialised into using English in their childhood. This difference could imply that Finnish teachers rely on the norms and conventions that they have been taught, while native teachers may simply trust their intuition. However, despite some more lenient tendencies among Finnish teachers, no systematic pattern arises. Furthermore, differences also appear in the acceptance

of some forms between the British and American teachers, indicating that the differences are, indeed, individual. Furthermore, many of the Finns had long experience in both studying and teaching English, which may already have automatised their responses to the level of intuition. Thus, based on this study, no convincing justification arises for categorically questioning the ability of non-native teachers of English (cf. Section 2.2).

In this study, variation among teachers was not considered a negative phenomenon that should be overcome through harmonisation attempts (e.g. McNamara 2000). Instead, variation was considered an interesting phenomenon requiring further research. It now seems evident that one rater on the acceptability of language (see Section 3.3.2) is insufficient, as different norms, varieties and standards can be used as the foundations for such assessments. More raters are needed to both ensure fairness and increase generalisability. As this study has shown, changing the level of agreement in assessing students' responses radically affects which forms are considered 'correct'.

Even rigorous attempts at training and harmonisation do not fully standardise teachers as raters. To some extent, students are at the mercy of their teachers' knowledge of the extent of variation in English. For example, a number of teachers in this study rated the form *spilt* as inappropriate, despite the fact that it appears in a number of British English dictionaries and grammar books as a legitimate variant of *spilled*. The same applies to teachers' familiarity with verb forms used after specific phrases. However, if the purpose of testing is to make inferences about learners' abilities to perform in a language, raters do not need to be fully unanimous on isolated features of language. In case the raters are gatekeepers (McNamara 2000, 4), the situation is different: if the decision made on the basis of the test affects people's lives in potentially radical ways (e.g. Read 2015, 475), seeking consensus is more important. In high-stakes tests, employing several testers would be useful to cater to various interpretations of the student data.

While this test included as many as thirteen teachers, more research is needed to investigate the existence of a feasible limit to the number of raters needed for fair assessment. In this study, the crucial line seems to be somewhere between four and ten, as the move from 10 to 13 only causes small differences in the results. A high number of raters is needed to explore the limits of variability in assessment. As we saw in Section 3.1, different cultural settings may cause different approaches to errors and to 'Standard' English. Researching English teachers from various backgrounds is important to allow investigation of the existence of similarities and differences. Furthermore, the understanding teachers have of grammar and standard use varies

and is linked to deep-rooted beliefs regarding language and its use, which form a part of teachers' psychological and professional identity.

Many teachers in Western countries have shown reluctance to teach grammar. As discussed in Section 3.1.2, Watson (2012; 2015) found that many secondary school teachers in Great Britain have negative attitudes to teaching grammar, and Alho and Korhonen (2018) reported that Finnish teachers avoid using the term 'grammar'. A similar tendency can be witnessed in some of the teachers in my study, while others had more positive views of grammar. However, most of the teachers did not seem to focus on teaching grammar: many of them prioritised fluency and communication. Perhaps a wider understanding of grammar, with a focus on the emancipatory power it gives for making informed choices, as argued by Larsen-Freeman (2002) and Celce-Murcia (2002), would help teachers appreciate grammar.

10.4 Implications

Important changes have occurred in both the Finnish and international school curricula over the past few decades: the emphasis has shifted from focusing on grammar to focusing on communication by encouraging students to use the skills they have, without worrying unnecessarily about grammatical accuracy (see Section 5.3). This study found no difference between older and younger students, despite the fact that some had studied according to the 1985 and others according to the 1994 curriculum. These curricula are very different in the level of precision and focus on particular aspects of grammar (see Section 5.3). The results of this study mirror those of Meriläinen (2010a; 2010b), who argues that changes in curricula have not affected Finns' skills in grammar.

In this light, it seems that "today's communicative based language teaching produces learners with a readiness to communicate in the foreign language but with relatively weak knowledge of its grammatical norms" (Meriläinen 2010b, 62): more focus on communicativeness has boosted learners' lexical skills but has not improved their accuracy. There may be no need to address the potential erroneous structures in Finnish students' production if the students are communicative (Meriläinen 2010b, 61) and intelligible (Thornbury 1999) and if they aim at fluency. However, if the students hope to pursue a career where accuracy in writing is key to professional credibility, such as an academic career, focusing on accuracy is more important (Meriläinen 2010a; 2010b). An emphasis on communicative language teaching "need not be at

the expense of attention to the rules of grammar, however. Relaxing on accuracy simply acknowledges the fact that the rules of grammar take a long time to establish themselves, and that, in the meantime, the learners' wish to communicate should not be needlessly frustrated" (Thornbury 1999, 23). Thus, focusing on fluency does not exclude focusing on accuracy. Although some aspects of grammar may be difficult to teach (Nassaji and Fotos 2011, 136), trying is always worthwhile.

Based on my findings, I can join Ellis and Barkhuizen (2005; see also Section 2.3.2) and Thewissen (2013, 94) in arguing that language learning is not a linear process and language performance is not consistent. Some learners occasionally first provided the expected form but later failed to do so, or "learned" to provide the standard form later in the test. The fluctuation was particularly common with spelling and the command of the progressive and past perfect, and students were more confident with more frequently occurring verbs. Inconsistency was also found among the most proficient students. How focused students were on the task at hand may explain differences in their consistency, particularly concerning tense discrepancy within the same sentence. As shown by Tabatabaei and Dehghani (2012), students may not be internally consistent over time, either.

The genre that was used for the test is not very common, as it contained a narrative in the form of monologues and a dialogue in written form but exemplifying spoken language. The genre possibly affected some students' decisions, and deciding on the level of formality to use in such a context can have been difficult for students. This study cannot answer the questions of how aware the students were of the requirements of style in different genres and whether they could make informed decisions regarding the provision of different forms in different contexts. Another question that remains open is how aware students were of the effects of the choice they made for one slot on the choices they made in subsequent slots. For example, if a student changes from the past tense to the present in the same sentence, is it a consciously made decision to allow reaching a particular (perhaps stylistic) effect, or is the student unaware of the conventional limitations of tense combinations within a sentence? Overall, more attention to accuracy would benefit students in the decisions they make in tense and aspect, as suggested by Hinkel (2002; 2004; 2011).

This study has shown that there is significant variation among students in their ability to provide the expected verb form. Although being both fluent and able to communicate with whatever skills you have is important, a proper command of verb forms is required in both professional

and academic writing. Students graduating from a university need to be more accurate to succeed in various duties at work. As indicated by Mauranen, Hynninen and Ranta (2010), academic writing needs more attention: everyone, including native speakers, need to learn the conventions and expectations of formal prose. Students need to become more aware of the requirements of genre: improved skills in the lexicon are not sufficient. Raising greater awareness of accuracy is critical to help the students who scored poorly, to allow them to improve their skills with a greater focus on the power of choice provided by a good command of grammar. Being communicative is not simply being fluent: accuracy plays a role in the intelligibility of the message as well.

This study has revealed that students with mark 8 from school or *magna cum laude approbatur* from the matriculation examination can either be very proficient or score poorly. While the top two marks fairly reliably predicted success in the test, mark 8 and *magna cum laude approbatur* already exhibited much variation, and some students with lower marks actually scored better. A subsequent conclusion may be that the top marks indicate skills in accuracy but that marks from 8 and *magna cum laude approbatur* down may also reflect other criteria. While rewarding students for fluency and other merits in their use of English is important, more harmonisation attempts may be necessary.

Based on this study, applying a simple dichotomy in assessing students' provision of verb forms is not particularly successful, because it may ignore some features present in particular instances. For example, the acceptability of using contractions in a particular context depends on the expectations associated with the genre present in the context. In this study, deciding on such acceptability became problematic, as the test represented written language but was based on informal speaking contexts. However, a possible model was available in the test, as the text included contracted forms in sentences without any slot to fill in. The fact that some teachers occasionally accepted contracted forms but dismissed their use in other slots compromises their ability to systematically assess students' contribution. Despite the difficulty in deciding what conventions the genre used in the text should follow and whether contractions should be allowed in such contexts, intra-rater consistency is important. Although some 'blind spots' in assessment are unavoidable, deciding on a set approach regarding contracted forms and using it systematically seems fairer to students. In contrast, it would be more feasible to judge the provision of tense and aspect slot by slot; here, a black-and-white assessment criterion would not support fairness. Moreover, this study has indicated that raters differ in their familiarity

with, for example, variety-specific forms. Perhaps teachers should also invest in awareness-raising regarding other varieties of English than their own.

As we have seen, a limited context, such as the test used here, already attracts high levels of variation, both in student responses and teacher reactions to such responses. Nevertheless, this variation is little compared to what teachers face in marking tests and student texts in course contexts. Exploring variation within a limited, partly controlled context enables assessment of the variables at play, but real-life situations involve accounting for more aspects of grammar and other factors, making assessment even more complex.

As we have seen, tightening the assessment criteria quickly leads to more student errors. The differences already increase significantly when moving from the lenient approach (one teacher's acceptance) to the moderate approach (four teachers' acceptance); the differences are even more radical when compared with the conservative level, which requires ten teachers' acceptance. In some cases, what is considered an error in learner language is considered innovative in native use, as argued by Edwards (2014) in Section 4.2. Assessing learner language can reflect teachers' tolerance of variation or tolerance of errors, which, again, reflects the understanding teachers have of grammar, norms and standards. Since there is variation in native and teacher use, perhaps more variation should be tolerated in learner English as well.

11 Conclusions

This chapter summarises the outcomes of this study. Section 11.1 outlines the contribution of this study to research on variation in English, language teacher cognition and teaching English as a second language. Section 11.2 discusses some limitations of the study, Section 11.3. suggests areas of further study, and Section 11.4 provides the final remarks.

11.1 The contribution of the study

This study has discussed manifestations of variation in English with a focus on both students and teachers. It provides further insights into the range of responses students display in their use of English verb forms and the range of variation in teachers' assessment of such verb forms. By exploring Finnish university students' proficiency in the use of English verb forms, this study deepens our understanding of the difficulties that Finnish students have in this domain of language use, a topic understudied so far. Furthermore, the study contributes to our understanding of the levels of English that Finnish university students have, whereas much of the previous research has focused on primary and secondary school levels or the population at large. Unlike many previous studies, this study is not contrastive but discusses variation arising from various factors.

Thus far, variation in teachers' assessment has also been understudied. This study contributes to research by providing insights into variation in teachers' acceptance of various verb forms by both native and non-native teachers of English. Here, variation is described and analysed, not simply condemned as a negative phenomenon that should be combatted with harmonisation attempts. This study finds that one rater on the acceptability of language phenomena is not sufficient, because different criteria can be used as bases for such assessments. More raters are needed to ensure the representation of varieties of English use, in order to avoid bias and to increase generalisability. Furthermore, more research is needed on rater variation.

This study has also explored whether native and non-native teachers differ in their rating of English verb forms and concludes that such differences are minimal, never systematic and typically idiosyncratic. In the same fashion, differences between British and American teachers

are negligible at the group level and more dependent on individual teacher characteristics. This suggests that there is no reason to consider Finnish teachers of English inferior to native teachers of English. The study also shows that in some cases, teachers cannot reach consensus on which verb form would best suit the context. The test that was used provided a realistic setting with a context and room for variation.

Knowledge about students' command of verb forms can provide valuable information to guide instruction in English as a second or foreign language and for lingua franca purposes. The results of this study can benefit tertiary education as well as at lower school levels. They also shed light on the extent of variation in learners and teachers of English and illustrate the differences between students majoring in English and other students. Furthermore, the results indicate which topics Finnish students would need to study more in order to gain better command of English verb forms. In particular, students would benefit from further instruction in the use of the progressive and the past perfect as well as in reported speech. Further training in the use of the passive and modal verbs would also be helpful.

The results suggest that even successful students' performance is not optimal if rated at a strict level: English Majors also provided forms that were considered inappropriate. This calls for a change of attitudes regarding grammar: if students are to become academically educated professionals, they can be expected to display accuracy in the choice of verb forms. Teachers should pay more attention to highlighting the crucial differences between English and Finnish at least to students whose goal is to gain mastery in English. For students who seek communicative competence but not full accuracy, such perfection may not be necessary. Ultimately, however, just as there are situations where fluency is expected, there is a time and place for accuracy, particularly at the university level. Communicative competence need not exclude accuracy. Some aspects of grammar are perhaps difficult to teach, but this should not prevent us from trying to do so, obviously acknowledging the fact that teaching does not equal learning (Ellis 2006).

The results suggest that increasing exposure to and use of English during visits to English-speaking countries and the consumption of media may promote students' fluency but not necessarily their accuracy. It seems that accuracy is best attained through instruction, whereas mere exposure, even when frequent, does not seem to provide the skills required for success. It may be useful to distinguish between instruction for lingua franca purposes, with a main focus

on fluency, and for more demanding academic purposes, with a focus on both fluency and accuracy.

While previous research has not sufficiently focused on the fact that raters, researchers and teachers can be biased in their evaluation of students' use of English, we need to acknowledge that one's own intuition may be misleading in some areas of grammar. We need greater awareness of the fact that although there may be consensus on the fact that a particular form is erroneous in a given context, there may be no consensus as to what the correct form would be, or there may be several correct or possible forms to choose from. It is important to remain aware of the limitations of one's own intuition with regard to normative language use.

11.2 Limitations

There are some issues with the test setting, the participants, the age of the data and the processing of the data that can limit the generalisability, reliability and validity of this study (e.g. Mackey and Gass 2005; Salkind 2006; 2008). In listing these limitations, I follow Borg's (2015) advice and attempt to be both detailed and (self-)critical. I begin the discussion with limitations related to participant conduct and issues that arose in data collection and coding. Finally, I discuss potential limitations caused by the test that was used and the fact that the data is from the early 21st century.

Teachers can potentially have had acquiescence bias or self-deception bias (Borg 2015; De Costa 2015; Wagner 2015), which would make them behave differently from what they actually believe because of attempting to adhere to an (assumed) ideal response or by attempting to respond in the way they assumed they were supposed to respond. Nevertheless, since the purpose of this study was to explore the extent of variation, not to find inter-rater agreement or a particular approach to strictness or leniency, such biases, if they exist, do not compromise any student results as such. They can, of course, partially distort the level of strictness or leniency indicated by the data. However, it is difficult to know whether the teachers would be more likely to pretend to be more strict or more lenient than they actually are, since no prestige difference has been reported between the two approaches. Therefore, such biases are not particularly likely to occur in this study, although teachers were probably tempted to consider what the expected answers were and to contemplate on how to react to them.

Some teachers' responses potentially began to display fatigue, exhaustion or boredom effects (e.g. Field 2013, 18; Mackey and Gass 2005; Wagner 2015), since the list of verb forms was quite exhausting. This could mean that the teachers may have marked more forms with 'inappropriate' than they would if they had been fully focused, and that they perhaps ceased to search for potential 'acceptable' and 'questionable' answers. However, on the whole, the full scale was used until the end, although some individuals may have changed their approach. Some teachers responded over the course of several days, which can also mean that they were more focused at the end of the list than in the middle, or vice versa. In contrast, most students reported having completed the test in 15 to 30 minutes. Given that the test is a story that is more entertaining than working on isolated sentences, fatigue is not likely to constitute a problem for students, although a few students did not provide an answer to the last items at all (for the effect of time in completing questionnaires, see e.g. Dörnyei and Csizér 2012, 78; Wagner 2015).

Section 6.5 discussed the reasons that caused the student and teacher lists to slightly differ from one another (see Appendix 6). While it is unfortunate that the lists are not identical, the effect of this on the results is small and only concerns a low percentage of responses. Since the differences concern the entire population and no student in particular either suffers or benefits from this, the significance can be considered minimal.

The test chosen for the study examines the use of verb forms, but their distribution is not even: some tenses and aspects are expected more often than others. However, the test consists of a story, and it is fairly difficult to create a story that would evenly measure all possible forms. Despite the main focus on past forms, the test constitutes a logical entity: the slots are in relation to one another, and what is provided in one slot affects the next one. Furthermore, as the overall aim was not to investigate any particular phenomenon in grammar but to explore trends in variation, the test in its present form serves this purpose well. Thus, the uneven distribution of forms does not compromise validity (e.g. Gass 2015; Phakiti 2015, 31; Révész 2012).

The study was conducted at three universities in the metropolitan area of Helsinki. It is possible that students elsewhere in Finland would have provided different answers and that students from other faculties at the University of Helsinki would have answered differently. As previous studies have found that Finns in southern cities outperform those in Northern and Eastern Finland (see Section 5.4.1), it is not known whether the results can be generalised to all university students in Finland. However, the size of the student population was large (319 students), which provides some grounds to assume at least limited generalisability.

Student data was collected in 2003-2004, which means that the knowledge gained from this study is partly becoming old. However, the wealth of data provided here yields an excellent point of comparison with more recent data. It would be easy to replicate the study to investigate whether students who have been brought up in an even more international context, surrounded by English from their early childhood, respond differently and whether their skills in English verb use have improved.

Finally, there may be some limitations with the study due to the fact that, as Pallant (2013, 4) puts it, "[p]eople are notoriously unreliable" and "don't fill out questionnaires properly", which not only involves this study but a vast number of others as well. Some students did not always respond to every slot, some teachers did not always mark every form provided, and some people's handwriting was difficult to interpret. Nevertheless, these limitations affect this study no more than any study that involves human behaviour (cf. McKay 2006, 36).

11.3 Further research

Although this study discovered many interesting features in variation among students and teachers, it also opens many possibilities for further research. More data already exist on the participating students' self-assessment of their success in the test and their experiences from learning English at school. Some students were also interviewed on their decisions to provide certain specific responses in the test and on their memories from English lessons at school (see Section 6.1). Such data will be used in future research to complement this study. Combining the results from this study with more qualitative results will provide enriched data on the effect of schooltime experiences and memories on student proficiency. Case studies or student profiles can also be drawn on the basis of the present data. Furthermore, as explained in Section 11.2, it would be useful to replicate the study in order to gain information on whether the results are relevant today as well.

Some ideas for further study arise from the fact that students were not equally successful in using the progressive as they were with the simple verb forms. It would be useful to study whether any patterns could be identified in the extent to which students can provide the progressive. A longitudinal study with the same participants would also help us examine whether students and teachers are consistent over time (cf. Tabatabaei and Dehghani 2012) in their provision and assessment of verb forms. In addition, while this study focused on variation

in verb forms, other grammatical structures could also be researched from the perspective of variation.

This study was conducted in the metropolitan area of Helsinki, and with students from a limited number of universities and faculties taking a course in English. Thus, students who had successfully applied for exemption from the course or students studying in other faculties might provide different answers. Similarly, results from students at other universities in Finland and in other countries might be different. Moreover, all the participants spoke Finnish or Swedish as their first language; students with other first languages might produce different results. All of these topics would merit further study. Future studies could also investigate whether students with L1 Finnish and L1 Swedish have higher or lower skills than do university students in other European countries.

It would also be interesting to study whether changes in school curricula result in changes in students' fluency and to compare students' progress in accuracy and fluency. Both Meriläinen (2010a; 2010b) and this study suggest that students' accuracy has not improved following the change from the 1985 to the 1994 curriculum. Therefore, it would be important to examine how the contemporary generation, with English being such an important part of their daily lives, succeeds in accuracy, and whether the increasing use of and exposure to English predominantly affect students' fluency and lexical skills. As English is now becoming a second rather than a foreign language in Finland, it is important to research what effects such changes have on Finns' language skills. Furthermore, it is worth investigating whether increased fluency increases variation in accuracy.

Regarding teacher variation, while extreme differences of opinion (i.e. ranging from the best to inappropriate) in rating the acceptability of students' responses were limited to a small number of slots, they still occurred to a surprisingly high degree. Unfortunately, the scope of the study prevented the use of interviews in the present study, but it would be useful to explore this further by interviewing teachers about their assessment to see if the differences can be explained in some way or whether they would be evened out after a discussion with other teachers. Furthermore, interview data could be compared to teachers' assessment data, and case studies and teacher profiles could also be created. It would also be useful to study variation among a greater number of teachers, as this study only included thirteen teachers. It is possible that a larger sample, including more teachers who have different first languages, would provide different results.

In this study, more men than women were strict in their evaluation of students' responses, and the most experienced teachers were neither lenient nor strict. It would be interesting to research whether this is coincidental or whether increasing experience makes you neither strict nor lenient but something between the two extremes. Furthermore, it would be worth exploring variation arising from the use of naïve informants (i.e. non-teachers) rating the acceptability of student responses.

11.4 Final remarks

Variation is extensive, pervasive and unavoidable. This does not mean that we should abandon attempts to harmonise teachers' behaviour in testing situations, but it does mean that some variation will always remain and that we should not worry about it too much. Variation is not a negative feature, as linguistic phenomena typically involve choice. Linguistic variation can be seen as an element that enriches life: language learners should be encouraged to try different ways of expression and be informed about the choices available. Unconventional forms are not always errors; at times, they may be a sign of new trends in a language. Furthermore, given the expansion of the use of English in the world, it only seems fair that L2 and ELF speakers could be part of the innovations and change that occur. Knowledge of English grammar allows learners to have more choice and control over how to formulate what they want to say or write, and focus on both fluency and accuracy enables students to make more informed choices in their communication.

Researchers in second language acquisition, in variation studies and in teacher assessment need to acknowledge the fact that teachers', raters' and researchers' reliance on intuition alone can lead to unjustified findings and ill-informed decisions. They need to pay more attention to the fact that language assessment cannot always find reliable norms to follow and that different raters' opinions may deviate from one another in radical ways. In order to treat learners fairly, testers and researchers should acknowledge that variation belongs to language use and that sometimes competing interpretations in the same context are possible: the expected alternative is not always the only acceptable alternative. Furthermore, teachers should be aware of the variability in their own assessment and the variation that is inherent in students' language use.

This study has shown that proficiency and accuracy are not simple concepts to define. Whether a particular student is considered accurate depends on who rates their performance and on how

many raters there are. Similarly, the concept of norm is problematic, as different teachers seem to follow different norms as their criteria for successful response. This implies that more awareness-raising is required among both students and teachers regarding the relation of accuracy and proficiency due to the fact that different applications of norms result in different assessments of accuracy and proficiency. This entails that variation persists and needs to be acknowledged in teacher training and that students need to become aware of the variation that exists in the use of grammatical items in English. Thus, some further attention to grammar would provide students with more options to express their message both accurately, proficiently, meaningfully and intelligibly.

Bibliography

Alanen, Riikka. 1997. *Grammaticality Judgments and Reaction Time Measurement: A Tool for Analyzing the Use of Second Language Knowledge*. Jyväskylä: University of Jyväskylä.

Alho, Irja and Korhonen, Riitta. 2018. Jälkiä häivytetystä kieliopista – opettajien näkemyksiä kieliopin opettamisesta. *Kieli, koulutus ja yhteiskunta*, 9 (3). Available at: www.kieli verkosto.fi/fi/journals/kieli-koulutus-ja-yhteiskunta-toukokuu-2018/jalkia-haivytetysta-kieliopista-opettajien-nakemyksia-kieliopin-opettamisesta (accessed on 23 June 2018).

Ambridge, Ben. 2010. Children's Judgments of Regular and Irregular Novel Past-Tense Forms: New Data on the English Past-Tense Debate. *Developmental Psychology*, 46 (6), 1497-1504.

Andrews, Stephen. 2007. *Teacher Language Awareness*. Cambridge: Cambridge University Press.

Bader, Markus and Häussler, Jana. 2010. Toward a Model of Grammaticality Judgments. *Journal of Linguistics*, 46 (2), 273-330.

Baralt, Melissa. 2012. Coding Qualitative Data. In: *Research Methods in Second Language Acquisition. A Practical Guide*, ed. by Mackey, Alison and Gass, Susan M. Chichester: Wiley-Blackwell, 222-244.

Barcelos, Ana Maria Ferreira and Kalaja, Paula. 2013. Beliefs in Second Language Acquisition: Teacher. In: *The Encyclopedia of Applied Linguistics*, ed. by Chapelle, Carol A. Available at: onlinelibrary.wiley.com/doi/pdf/10.1002/9781405198431.wbeal0083 (accessed on 5 July 2018).

Bardovi-Harlig, Kathleen. 1998. Narrative Structure and Lexical Aspect: Conspiring Factors in Second Language Acquisition of Tense-Aspect Morphology. *Studies in Second Language Acquisition*, 20, 471-508.

Bardovi-Harlig, Kathleen and Dörnyei, Zoltán. 1998. Do Language Learners Recognize Pragmatic Violations? Pragmatic Versus Grammatical Awareness in Instructed L2 Learning. *TESOL Quarterly*, 32 (2), 233-259.

Bayley, Robert and Langman, Juliet. 2004. Variation in the Group and the Individual: Evidence from Second Language Acquisition. *International Review of Applied Linguistics in Language Teaching*, 42 (2), 303-318.

Benati, Alessandro G. and Angelovska, Tanja. 2016. *Second Language Acquisition. A Theoretical Introduction to Real World Applications*. London: Bloomsbury.

Benzie, Helen Joy. 2010. Graduating as a 'Native Speaker': International Students and English Language Proficiency in Higher Education. *Higher Education Research and Development*, 29 (4), 447-459.

Biber, Douglas; Johansson, Stig; Leech, Geoffrey; Conrad, Susan and Finegan, Edward. 1999. *Longman Grammar of Spoken and Written English*. Harlow: Longman.

Birrell, Bob. 2006. Implications of Low English Standards Among Overseas Students at Australian Universities. *People and Place*, 14 (4), 53-64.

Boas, Hans C. 2011. Coercion and Leaking Argument Structures in Construction Grammar. *Linguistics*, 49 (6), 1271-1303.

Borg, Simon. 2006. *Teacher Cognition and Language Education: Research and Practice*. London: Continuum.

Borg, Simon. 2015. Researching Teachers' Beliefs. In: *Research Methods in Applied Linguistics: A Practical Resource*, ed. by Paltridge, Brian and Phakiti, Aek. London: Bloomsbury, 487-504.

Brown, H. Douglas. 2007a. *Principles of Language Learning and Teaching*. 5th ed. White Plains: Pearson.

Brown, H. Douglas. 2007b. *Teaching by Principles: An Interactive Approach to Language Pedagogy*. 3rd ed. White Plains: Pearson.

Brown, James Dean. 2001. *Using Surveys in Language Programs*. Cambridge: Cambridge University Press.

Brown, James Dean. 2011. Quantitative Research in Second Language Studies. In: *Handbook of Research in Second Language Teaching and Learning: Volume II*, ed. by Hinkel, Eli. New York: Routledge, 190-206.

Brown, James Dean and Rodgers, Theodore S. 2002. *Doing Second Language Research*. Oxford: Oxford University Press.

Brown, Steven and Larson-Hall, Jenifer. 2012. *Second Language Acquisition Myths. Applying Second Language Research to Classroom Teaching*. Ann Arbor: University of Michigan Press.

Butcher, Carmen Acevedo. 2005. The Case Against the 'Native Speaker'. *English Today*, 21 (2), 13-24.

Casanave, Christine Pearson. 2015. Case Studies. In: *Research Methods in Applied Linguistics: A Practical Resource*, ed. by Paltridge, Brian and Phakiti, Aek. London: Bloomsbury, 119-135.

CEFR = Council of Europe. 2001. *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*. Cambridge: Cambridge University Press. Also available at: www.coe.int/t/dg4/linguistic/Source/Framework_EN.pdf (accessed on 29 November 2016).

Celce-Murcia, Marianne. 2002. Why It Makes Sense to Teach Grammar in Context and Through Discourse. In: *New Perspectives on Grammar Teaching in Second Language Classrooms*, ed. by Hinkel, Eli and Fotos, Sandra. Mahwah: Lawrence Erlbaum, 119-133.

Celce-Murcia, Marianne and Larsen-Freeman, Diane (with Williams, Howard). 1999. *The Grammar Book. An ESL/EFL Teacher's Course*. 2nd ed. Boston: Heinle and Heinle.

Cohen, Louis; Manion, Lawrence and Morrison, Keith. 2011. *Research Methods in Education*. 7th ed. London: Routledge.

Collins, Laura. 2004. The Particulars on Universals: A Comparison of the Acquisition of Tense-Aspect Morphology Among Japanese- and French-Speaking Learners of English. *The Canadian Modern Language Review*, 61 (2), 251-274.

Collins, Laura. 2005. Accessing Second Language Learners' Understanding of Temporal Morphology. *Language Awareness*, 14 (4), 207-220.

Collins, Peter. 2008. The Progressive Aspect in World Englishes: A Corpus-Based Study. *Australian Journal of Linguistics*, 28 (2), 225-249.

Cook, Vivian. 2001. *Second Language Learning and Language Teaching*. 3rd ed. London: Arnold.

Cook, Vivian. 2011. Teaching English as a Foreign Language in Europe. In: *Handbook of Research in Second Language Teaching and Learning: Volume II*, ed. by Hinkel, Eli. New York: Routledge, 140-154.

Cook, Vivian. 2016. Transfer and the Relationships Between the Languages of Multi-Competence. In: *Crosslinguistic Influence in Second Language Acquisition*, ed. by Alonso Alonso, Rosa. Bristol: Multilingual Matters, 24-37.

Crossley, Scott A. and McNamara, Danielle S. 2011. Shared Features of L2 Writing: Intergroup Homogeneity and Text Classification. *Journal of Second Language Writing*, 20, 271-285.

Crystal, David. 2003. *The Cambridge Encyclopedia of the English Language*. 2nd ed. Cambridge: Cambridge University Press.

Dąbrowska, Ewa. 2010. Naive v. Expert Intuitions: An Empirical Study of Acceptability Judgments. *The Linguistic Review*, 27, 1-23.

Davies, William D. and Kaplan, Tamar I. 1998. Native Speaker vs. L2 Learner Grammaticality Judgements. *Applied Linguistics*, 19 (2), 183-203.

De Costa, Peter. 2015. Ethics and Applied Linguistics Research. In: *Research Methods in Applied Linguistics. A Practical Resource*, ed. by Paltridge, Brian and Phakiti, Aek. London: Bloomsbury, 245-257.

DeKeyser, Robert M. 2005. What Makes Learning Second-Language Grammar Difficult? A Review of Issues. *Language Learning*, 55 (S1), 1-25.

Demont-Heinrich, Christof. 2010. Linguistically Privileged and Cursed? American University Students and the Global Hegemony of English. *World Englishes*, 29 (2), 281-298.

Dewaele, Jean-Marc. 2013. Learner-Internal Psychological Factors. In: *The Cambridge Handbook of Second Language Acquisition*, ed. by Herschensohn, Julia and Young-Scholten, Martha. Cambridge: Cambridge University Press, 159-179.

Dinsmore, Thomas H. 2006. Principles, Parameters, and SLA. A Retrospective Meta-Analytic Investigation into Adult L2 Learners' Access to Universal Grammar. In: *Synthesizing Research on Language Learning and Teaching*, ed. by Norris, John M. and Ortega, Lourdes. Amsterdam: John Benjamins, 53-90.

Duff, Patricia A. 2012. How to Carry Out Case Study Research. In: *Research Methods in Second Language Acquisition: A Practical Guide*, ed. by Mackey, Alison and Gass, Susan M. Chichester: Wiley-Blackwell, 95-116.

Dörnyei, Zoltán. 2007. *Research Methods in Applied Linguistics: Quantitative, Qualitative, and Mixed Methodologies*. Oxford: Oxford University Press.

Dörnyei, Zoltán and Csizér, Kata. 2012. How to Design and Analyze Surveys in Second Language Acquisition Research. In: *Research Methods in Second Language Acquisition: A Practical Guide*, ed. by Mackey, Alison and Gass, Susan M. Chichester: Wiley-Blackwell, 74-94.

Dörnyei, Zoltán and Ryan, Stephen. 2015. *The Psychology of the Language Learner Revisited*. New York: Routledge.

Dörnyei, Zoltán and Skehan, Peter. 2003. Individual Differences in Second Language Learning. In: *The Handbook of Second Language Acquisition*, ed. by Doughty, Catherine J. and Long, Michael H. Malden: Blackwell, 589-630.

Edwards, Alison. 2014. The Progressive Aspect in the Netherlands and the ESL/EFL Continuum. *World Englishes*, 33 (2), 173-194.

EF (Education First). 2015. EF EPI: *EF English Proficiency Index*. Available at: mediakey1.ef.com/__/~/media/centralefcom/epi/downloads/full-reports/v5/ef-epi-2015-english.pdf (accessed on 31 July 2016).

Ellis, Rod. 2002a. Grammar Teaching – Practice or Consciousness-Raising? In: *Methodology in Language Teaching: An Anthology of Current Practice*, ed. by Richards, Jack C. and Renandya, Willy A. Cambridge: Cambridge University Press, 167-174.

Ellis, Rod. 2002b. The Place of Grammar Instruction in the Second/Foreign Language Curriculum. In: *New Perspectives on Grammar Teaching in Second Language Classrooms*, ed. by Hinkel, Eli and Fotos, Sandra. Mahwah: Lawrence Erlbaum, 17-34.

Ellis, Rod. 2006. Current Issues in the Teaching of Grammar: An SLA Perspective. *TESOL Quarterly*, 40 (1), 83-107.

Ellis, Rod and Barkhuizen, Gary. 2005. *Analysing Learner Language*. Oxford: Oxford University Press.

Ellis, Rod; Sheen, Younghee; Murakami, Mihoko and Takashima, Hide. 2008. The Effects of Focused and Unfocused Written Corrective Feedback in an English as a Foreign Language Context. *System*, 36, 353-371.

European Commission. 2012. *Special Eurobarometer 286: Europeans and their languages: Report*. Available at: ec.europa.eu/public_opinion/archives/ebs/ebs_386_en.pdf (accessed on 31 July 2016).

Field, Andy. 2013. *Discovering Statistics Using IBM SPSS Statistics: And Sex and Drugs and Rock 'n' Roll*. 4th ed. Los Angeles: SAGE.

Finegan, Edward. 2004. *Language: Its Structure and Use*. 4th ed. Boston: Wadsworth.

Finnish National Agency for Education. n.d. *Education System*. Available at: www.oph.fi/english/education_system (accessed on 29 November 2016).

Foley, Claire and Flynn, Suzanne. 2013. The Role of the Native Language. In: *The Cambridge Handbook of Second Language Acquisition*, ed. by Herschensohn, Julia and Young-Scholten, Martha. Cambridge: Cambridge University Press, 97-113.

Fotos, Sandra. 2002. Structure-Based Interactive Tasks for the EFL Grammar Learner. In: *New Perspectives on Grammar Teaching in Second Language Classrooms*, ed. by Hinkel, Eli and Fotos, Sandra. Mahwah: Lawrence Erlbaum, 135-154.

Foucart, Alice and Frenck-Mestre, Cheryl. 2013. Language Processing. In: *The Cambridge Handbook of Second Language Acquisition*, ed. by Herschensohn, Julia and Young-Scholten, Martha. Cambridge: Cambridge University Press, 394-416.

Friedrich, Patricia. 2010. Comment on Demont-Heinrich's "Linguistically Privileged and Cursed? American University Students and the Global Hegemony of English". *World Englishes*, 29 (2), 299-301.

Fulcher, Glenn. 2016. Standards and Frameworks. In: *Handbook of Second Language Assessment*, ed. by Tsagari, Dina and Banerjee, Jayanti. Boston: De Gruyter, 29-44.

Gass, Susan. 2015. Experimental Research. In: *Research Methods in Applied Linguistics: A Practical Resource*, ed. by Paltridge, Brian and Phakiti, Aek. London: Bloomsbury, 101-117.

Gass, Susan M. and Selinker, Larry. 2001. *Second Language Acquisition: An Introductory Course*. 2nd ed. Mahwah: Lawrence Erlbaum.

Gibson, Edward and Fedorenko, Evelina. 2013. The Need for Quantitative Methods in Syntax and Semantics Research. *Language and Cognitive Processes*, 28 (1/2), 88-124.

Goldschneider, Jennifer M. and DeKeyser, Robert M. 2001. Explaining the "Natural Order of L2 Morpheme Acquisition" in English: A Meta-analysis of Multiple Determinants. *Language Learning*, 51 (1), 1-50.

Gorard, Stephen. 2010. Research Design, as Independent of Methods. In: *SAGE Handbook of Mixed Methods in Social & Behavioral Research*. 2nd ed., ed. by Tashakkori, Abbas and Teddlie, Charles. Los Angeles: SAGE, 237-251.

Gut, Ulrike and Fuchs, Robert. 2013. Progressive Aspect in Nigerian English. *Journal of English Linguistics*, 41 (3), 243-267.

Hackert, Stephanie. 2009. Linguistic Nationalism and the Emergence of the English Native Speaker. *European Journal of English Studies*, 13 (3), 305-317.

Hakulinen, Auli; Vilkuna, Maria; Korhonen, Riitta; Koivisto, Vesa: Heinonen, Tarja Riitta and Alho, Irja (eds.). 2004. *Iso suomen kielioppi*. Helsinki: Suomalaisen kirjallisuuden seura.

Han, ZhaoHong. 2011. Fossilization – A Classic Concern of SLA Research. In: *The Routledge Handbook of Second Language Acquisition*, ed. by Gass, Susan and Mackey, Alison. London: Routledge, 476-490.

Hashemi, Louise and Murphy, Raymond. 1995. *English Grammar in Use Supplementary Exercises: With Answers*. Cambridge: Cambridge University Press.

Hawkins, Roger. 2007. The Nativist Perspective on Second Language Acquisition. *Lingua*, 118, 465-477.

Hawkins, Roger and Casillas, Gabriela. 2008. Explaining Frequency of Verb Morphology in Early L2 Speech. *Lingua*, 118, 595-612.

Hedge, Tricia. 2000. *Teaching and Learning in the Language Classroom*. Oxford: Oxford University Press.

Herschensohn, Julia. 2013. Age-Related Effects. In: *The Cambridge Handbook of Second Language Acquisition*, ed. by Herschensohn, Julia and Young-Scholten, Martha. Cambridge: Cambridge University Press, 317-337.

Hewings, Martin. 2005. *Advanced Grammar in Use: A Self-Study Reference and Practice Book for Advanced Learners of English: With Answers*. 2nd ed. Cambridge: Cambridge University Press.

Hinkel, Eli. 2002. Teaching Grammar in Writing Classes: Tenses and Cohesion. In: *New Perspectives on Grammar Teaching in Second Language Classrooms*, ed. by Hinkel, Eli and Fotos, Sandra. Mahwah: Lawrence Erlbaum, 181-198.

Hinkel, Eli. 2004. Tense, Aspect and the Passive Voice in L1 and L2 Academic Texts. *Language Teaching Research*, 8 (1), 5-29.

Hinkel, Eli. 2011. What Research on Second Language Writing Tells Us and What It Doesn't. In: *Handbook of Research in Second Language Teaching and Learning: Volume II*, ed. by Hinkel, Eli. New York: Routledge, 523-538.

Hinkel, Eli and Fotos, Sandra. 2002. From Theory to Practice: A Teacher's View. In: *New Perspectives on Grammar Teaching in Second Language Classrooms*, ed. by Hinkel, Eli and Fotos, Sandra. Mahwah: Lawrence Erlbaum, 1-12.

Holmes, Philip and Hinchliffe, Ian. 2013. *Swedish. A Comprehensive Grammar*. 3rd ed. London: Routledge.

Hos, Rabia and Kekec, Mustafa. 2014. The Mismatch Between Non-Native English as a Foreign Language (EFL) Teachers' Grammar Beliefs and Classroom Practices. *Journal of Language Testing and Research*, 5 (1), 80-87.

Huhta, Ari; Alanen, Riikka; Tarnanen, Mirja; Martin, Maisa and Hirvelä, Tuija. 2014. Assessing Learners' Writing Skills in a SLA Study: Validating the Rating Process across Tasks, Scales and Languages. *Language Testing* 31 (3), 307-328.

Hyland, Ken and Anan, Eri. 2006. Teachers' Perceptions of Error: The Effects of First Language and Experience. *System*, 34, 509-519.

Ionin, Tania. 2012. Formal Theory-Based Methodologies. In: *Research Methods in Second Language Acquisition: A Practical Guide*, ed. by Mackey, Alison and Gass, Susan M. Chichester: Wiley-Blackwell, 30-52.

Ionin, Tania. 2013. Morphosyntax. In: *The Cambridge Handbook of Second Language Acquisition*, ed. by Herschensohn, Julia and Young-Scholten, Martha. Cambridge: Cambridge University Press, 505-528.

Jarvis, Scott. 2000. Methodological Rigor in the Study of Transfer: Identifying L1 Influence in the Interlanguage Lexicon. *Language Learning*, 50 (2), 245-309.

Jarvis, Scott. 2015. Influences of Previously Learned Languages on the Learning and Use of Additional Languages. In: *Content-Based Language Learning in Multilingual Educational Environments*, ed. by Juan-Garau, Maria and Salazar-Noruega, Joana. Cham: Springer, 69-86.

Jarvis, Scott and Odlin, Terence. 2000. Morphological Type, Spatial Reference, and Language Transfer. *Studies in Second Language Acquisition*, 22 (4), 535-556.

Jarvis, Scott and Pavlenko, Aneta. 2010. *Crosslinguistic Influence in Language and Cognition*. London: Routledge.

Jean, Gladys and Simard, Daphnée. 2011. Grammar Teaching and Learning in L2: Necessary, but Boring? *Foreign Language Annals*, 44 (3), 467-494.

Jenkins, Jennifer. 2009. English as a Lingua Franca: Interpretations and Attitudes. *World Englishes*, 28 (2), 200-207.

Jenkins, Jennifer; Cogo, Alessia and Dewey, Martin. 2011. Review of Developments in Research into English as a Lingua Franca. *Language Teaching*, 44 (3), 281-315.

Johnson, R. Burke and Onwuegbuzie, Anthony J. 2004. Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher* 33 (7), 14-26.

Kalaja, Paula. 2015a. 'Dreaming Is Believing': The Teaching of Foreign Languages as Envisioned by Student Teachers. In: *Beliefs, Agency and Identity in Foreign Language Learning and Teaching*, ed. by Kalaja, Paula et al. Basingstoke: Palgrave Macmillan, 124-146.

Kalaja, Paula. 2015b. Student Teachers' Beliefs About L1 and L2 Discursively Constructed: A Longitudinal Study of Interpretative Repertoires. In: *Beliefs, Agency and Identity in Foreign Language Learning and Teaching*, ed. by Kalaja, Paula et al. Basingstoke: Palgrave Macmillan, 97-123.

Karlsson, Fred. 2008. *Finnish. An Essential Grammar*. 2nd ed. London: Routledge.

Keck, Casey and Kim, YouJin. 2014. *Pedagogical Grammar*. Amsterdam: John Benjamins.

Kirkpatrick, Andy; Deterding, David and Wong, Jennie. 2008. The International Intelligibility of Hong Kong English. *World Englishes*, 27 (3), 359-377.

Klein, Elaine C. 1995. Second versus Third Language Acquisition: Is There a Difference? *Language Learning*, 45 (3), 419-465.

Kline, Rex B. 2004. *Beyond Significance Testing: Reforming Data Analysis Methods in Behavioral Research*. Washington: American Psychological Association.

Kondo-Brown, Kimi. 2002. A FACETS Analysis of Rater Bias in Measuring Japanese Second Language Writing Performance. *Language Testing*, 19 (1), 3-31.

Kouluhallitus. 1985. *Lukion opetussuunnitelman perusteet 1985*. Helsinki: Valtion painatuskeskus.

Laerd Statistics. 2013. Available at: statistics.laerd.com/ (accessed on 30 June 2016).

Lardiere, Donna. 2013. Native and Non-Nativelike Attainment. In: *The Cambridge Handbook of Second Language Acquisition*, ed. by Herschensohn, Julia and Young-Scholten, Martha. Cambridge: Cambridge University Press, 670-691.

Larsen-Freeman, Diane. 2002. The Grammar of Choice. In: *New Perspectives on Grammar Teaching in Second Language Classrooms*, ed. by Hinkel, Eli and Fotos, Sandra. Mahwah: Lawrence Erlbaum, 103-118.

Larsen-Freeman, Diane. 2009. Teaching and Testing Grammar. In: *The Handbook of Language Teaching*, ed. by Long, Michael H. and Doughty, Catherine J. Chichester: Blackwell, 518-542.

Leech, Geoggrey; Hundt, Marianne; Mair, Christian and Smith, Nicholas. 2009. *Change in Contemporary English. A Grammatical Study*. Cambridge: Cambridge University Press.

Lehtonen, Jaakko and Sajavaara, Kari. 1983. Acceptability and Ambiguity in Native and Second Language Message Processing. In: *Psycholinguistics and Foreign Language Learning: Papers from a Conference Held in Stockholm and Åbo, October 25-26, 1982*, ed. by Ringbom, Håkan. Åbo: Publications of the Research Institute of the Åbo Akademi Foundation, 101-125.

Leith, Dick (with revision by Jackson, Liz). 2007. English – Colonial to Postcolonial. In: *Changing English*, ed. by Graddol, David et al. Abingdon: The Open University, 117-148.

Leppänen, Sirpa and Nikula, Tarja. 2008. Johdanto. In: *Kolmas kotimainen: Lähikuvia englannin käytöstä Suomessa*, ed. by Leppänen, Sirpa et al. Helsinki: Suomalaisen kirjallisuuden seura, 9-40.

Leppänen, Sirpa; Pitkänen-Huhta, Anne; Nikula, Tarja; Kytölä, Samu; Törmäkangas, Timo; Nissinen, Kari; Kääntä, Leila; Räisänen, Tiina; Laitinen, Mikko; Koskela, Heidi, Lähdesmäki, Salla and Jousmäki, Henna. 2011. *National Survey on the English Language in Finland: Uses, Meanings and Attitudes.* Studies in Variation, Contacts and Change in English, vol. 5. Available at: www.helsinki.fi/varieng/journal/volumes/05 (accessed on 7 May 2017)

Lightbown, Patsy M. and Spada, Nina. 2006. *How Languages are Learned*. 3rd ed. Oxford: Oxford University Press.

Lindemann, Stephanie. 2002. Listening with an Attitude: A Model of Native-Speaker Comprehension of Non-Native Speakers in the United States. *Language in Society*, 31, 419-441.

Long, Michael H. 2003. Stabilization and Fossilization in Interlanguage Development. In: *The Handbook of Second Language Acquisition*, ed. by Doughty, Catherine J. and Long, Michael H. Malden: Blackwell, 487-535.

Luk, Zoe Pei-sui and Shirai, Yasuhiro. 2009. Is the Acquisition Order of Grammatical Morphemes Impervious to L1 Knowledge? Evidence from the Acquisition of Plural *–s*, Articles, and Possessive '*s*. *Language Learning*, 59 (4), 721-754.

Mackey, Alison and Gass, Susan M. 2005. *Second Language Research: Methodology and Design*. Mahwah: Lawrence Erlbaum.

Macmillan English Dictionary for Advanced Learners. 2002. Oxford: Macmillan.

Mandell, Paul B. 1999. On the Reliability of Grammaticality Judgement Tests in Second Language Acquisition Research. *Second Language Research*, 15 (1), 73-99.

Mauranen, Anna; Hynninen, Niina and Ranta, Elina. 2010. English as an Academic Lingua Franca: The ELFA Project. *English for Specific Purposes*, 29, 183-190.

McCarthy, Michael and Carter, Ronald. 2002. Ten Criteria for a Spoken Grammar. In: *New Perspectives on Grammar Teaching in Second Language Classrooms*, ed. by Hinkel, Eli and Fotos, Sandra. Mahwah: Lawrence Erlbaum, 51-75.

McDonald, Janet L. 2000. Grammaticality Judgments in a Second Language: Influences of Age of Acquisition and Native Language. *Applied Psycholinguistics*, 21, 395-423.

McKay, Sandra Lee. 2006. *Researching Second Language Classrooms*. New York: Routledge.

McNamara, Tim. 2000. *Language Testing*. Oxford: Oxford University Press.

Meriläinen, Lea. 2010a. *Language Transfer in the Written English of Finnish Students*: Publications of the University of Eastern Finland, Dissertations in Education, Humanities and Theology 9. Joensuu: University of Eastern Finland.

Meriläinen, Lea. 2010b. Syntactic Transfer in the Written English of Finnish Students: Persistent Grammar Errors or Acceptable Lingua Franca English? *Apples – Journal of Applied Language Studies*, 4 (1), 51-64.

Moreno, Sylvain; Bialystok, Ellen; Wodniecka, Zofia and Alain, Claude. 2010. Conflict Resolution in Sentence Processing by Bilinguals. *Journal of Neurolinguistics*, 23, 564-579.

Muñoz, Carmen and Gilabert, Roger. 2011. More Evidence Concerning the Aspect Hypothesis: The Acquisition of English Progressive Aspect by Catalan-Spanish Instructed Learners. *IRAL*, 49, 241-263.

Murakami, Akira and Alexopoulou, Theodora. 2016. L1 Influence on the Acquisition Order of English Grammatical Morphemes. *Studies in Second Language Acquisition*, 38, 365-401.

Murphy, Raymond. 2004. *English Grammar in Use: A Self-Study Reference and Practice Book for Intermediate Students of English: With Answers*. 3rd ed. Cambridge: Cambridge University Press.

Myles, Florence. 2013. Theoretical Approaches. In: *The Cambridge Handbook of Second Language Acquisition*, ed. by Herschensohn, Julia and Young-Scholten, Martha. Cambridge: Cambridge University Press, 46-70.

Nassaji, Hossein and Fotos, Sandra. 2011. *Teaching Grammar in Second Language Classrooms. Integrating Form-Focused Instruction in Communicative Context.* New York: Routledge.

Norris, John M. and Ortega, Lourdes. 2011. Assessing Learner Knowledge. In: *The Routledge Handbook of Second Language Acquisition*, ed. by Gass, Susan and Mackey, Alison. London: Routledge, 573-589.

Odlin, Terence. Cross-Linguistic Influence. In: *The Handbook of Second Language Acquisition*, ed. by Doughty, Catherine J. and Long, Michael H. Malden: Blackwell, 436-486.

Ojutkangas, Krista; Larjavaara, Meri; Miestamo, Matti and Ylikoski, Jussi. 2014. *Johdatus kielitieteeseen*. Helsinki: Sanoma Pro.

Oliver, Paul. 2010. *Understanding the Research Process*. Los Angeles: SAGE.

Opetushallitus. 1994. *Lukion opetusuunnitelman perusteet 1994*. Helsinki: Painatuskeskus.

Opetushallitus. 2003. *Lukion opetussuunnitelman perusteet 2003*. *Nuorille tarkoitetun lukiokoulutuksen opetussuunnitelman perusteet*. Vammala: Vammalan kirjapaino.

Opetushallitus. 2015. *Lukion opetussuunnitelman perusteet 2015*. Määräykset ja ohjeet 2015:48. Available at: www.oph.fi/download/172124_lukion_opetussuunnitelman_ perusteet_ 2015.pdf (accessed on 20 June 2016).

Ortega, Lourdes. 2009. *Understanding Second Language Acquisition*. London: Routledge.

Pallant, Julie. 2013. *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using IBM SPSS*. 5th ed. Maidenhead: McGraw Hill.

Palmberg, Rolf. 1977. Om finnars och finlandssvenskars engelska språkfel på olika nivåer. In: *Föredrag vid konferensen om kontrastiv lingvistik och felanalys, Stockholm & Åbo, 7-8 februari 1977 / Papers from the Conference on Contrastive Linguistics and Error Analysis, Stockholm & Åbo, 7-8 February 1977*, ed. by Palmberg, Rolf and Ringbom, Håkan. Åbo: Publications of the Research Institute of the Åbo Akademi foundation, 85-93.

Paltridge, Brian and Phakiti, Aek. 2015. *Developing a Research Project. In: Research Methods in Applied Linguistics: A Practical Resource*, ed. by Paltridge, Brian and Phakiti, Aek. London: Bloomsbury, 259-278.

Papageorgiou, Spiros. 2016. Aligning Language Assessments to Standards and Frameworks. In: *Handbook of Second Language Assessment*, ed. by Tsagari, Dina and Banerjee, Jayanti. Boston: De Gruyter, 327-340.

Pennington, Martha C. 2002. Grammar and Communication: New Directions in Theory and Practice. In: *New Perspectives on Grammar Teaching in Second Language Classrooms*, ed. by Hinkel, Eli and Fotos, Sandra. Mahwah: Lawrence Erlbaum, 77-98.

Pesola, Satu. 2002. *Kaikkihan englantia osaavat! Yliopisto-opiskelijoiden englannin kielen verbimuotojen hallinnasta*. Unpublished Bachelor-level thesis, Faculty of Education, University of Helsinki.

Phakiti, Aek. 2015. Quantitative Research and Analysis. In: *Research Methods in Applied Linguistics: A Practical Resource*, ed. by Paltridge, Brian and Phakiti, Aek. London: Bloomsbury, 27-49.

Phakiti, Aek and Paltridge, Brian. 2015. Approaches and Methods in Applied Linguistics Research. In: *Research Methods in Applied Linguistics: A Practical Resource*, ed. by Paltridge, Brian and Phakiti, Aek. London: Bloomsbury, 5-25.

Pienemann, Manfred. 2005. An Introduction to Processability Theory. In: *Cross-Linguistic Aspects of Processability Theory*, ed. by Pienemann, Manfred. Amsterdam: John Benjamins, 1-60.

Piispa, Minna. 2006. Kvantitatiivisen tutkimuksen eettiset lähtökohdat: Esimerkkinä naisiin kohdistuvan väkivallan kyselytutkimus. In: *Etiikkaa ihmistieteille*, ed. by Hallamaa, Jaana et al. Helsinki: Suomalaisen kirjallisuuden seura, 141-160.

Rasinger, Sebastian M. 2008. *Quantitative Research in Linguistics: An Introduction*. London: Bloomsbury.

Rautionaho, Paula. 2014. *Variation in the Progressive: A Corpus-Based Study into World Englishes*. Tampere: Tampere University Press.

Read, John. 2015. Researching Language Testing and Assessment. In: *Research Methods in Applied Linguistics: A Practical Resource*, ed. by Paltridge, Brian and Phakiti, Aek. London: Bloomsbury, 471-486.

Regan, Vera. 2004. The Relationship Between the Group and the Individual and the Acquisition of Native Speaker Variation Patterns: A Preliminary Study. *International Review of Applied Linguistics in Language Teaching*, 42 (2), 335-348.

Regan, Vera. 2013. Variation. In: *The Cambridge Handbook of Second Language Acquisition*, ed. by Herschensohn, Julia and Young-Scholten, Martha. Cambridge: Cambridge University Press, 272-291.

Révész, Andrea. 2012. Coding Second Language Data Validly and Reliably. In: *Research Methods in Second Language Acquisition: A Practical Guide*, ed. by Mackey, Alison and Gass, Susan M. Chichester: Wiley-Blackwell, 203-221.

Rhys, Martin (based on Thomas, Linda). 2007. Dialect Variation in English. In: *Changing English*, ed. by Graddol, David et al. Abingdon: The Open University, 189-221.

Richards, Keith. 2011. Case Study. In: *Handbook of Research in Second Language Teaching and Learning: Volume II*, ed. by Hinkel, Eli. New York: Routledge, 207-221.

Ringbom, Håkan. 1977. Spelling Errors and Foreign Language Learning Strategies. In: *Föredrag vid konferensen om kontrastiv lingvistik och felanalys, Stockholm & Åbo, 7-8 februari 1977 / Papers from the Conference on Contrastive Linguistics and Error Analysis, Stockholm & Åbo, 7-8 February 1977*, ed. by Palmberg, Rolf and Ringbom, Håkan. Åbo: Publications of the Research Institute of the Åbo Akademi foundation, 101-111.

Ringbom, Håkan. 1983. On the Distinctions of Item Learning vs. System Learning and Receptive Competence vs. Productive Competence in Relation to the Role of L1 in Foreign Language Learning. In: *Psycholinguistics and Foreign Language Learning: Papers from a Conference Held in Stockholm and Åbo, October 25-26, 1982*, ed. by Ringbom, Håkan. Åbo: Publications of the Research Institute of the Åbo Akademi Foundation, 163-173.

Ringbom, Håkan. 2016. Comprehension, Learning and Production of Foreign Languages: The Role of Transfer. In: *Crosslinguistic Influence in Second Language Acquisition*, ed. by Alonso Alonso, Rosa. Bristol: Multilingual Matters, 38-52.

Ringbom, Håkan and Jarvis, Scott. 2009. The Importance of Cross-Linguistic Similarity in Foreign Language Learning. In: *The Handbook of Language Teaching*, ed. by Long, Michael H. and Doughty, Catherine J. Chichester: Blackwell, 106-118.

Robenalt, Clarice and Goldberg, Adele E. 2016. Nonnative Speakers Do Not Take Competing Alternative Expressions Into Account the Way Native Speakers Do. *Language Learning*, 66 (1), 60-93.

Roberts, Leah and Liszka, Sarah Ann. 2013. Processing Tense/Aspect-Agreeement Violations On-Line in the Second Language: A Self-Paced Reading Study with French and German L2 Learners of English. *Second Language Research*, 29 (4), 413-439.

Romaine, Suzanne. 2003. Variation. In*: The Handbook of Second Language Acquisition*, ed. by Doughty, Catherine J. and Long, Michael H. Malden: Blackwell, 409-435.

Rothman, Jason, Cabrelli Amaro, Jennifer and de Bot, Kees. 2013. Third Language Acquisition. In: *The Cambridge Handbook of Second Language Acquisition*, ed. by Herschensohn, Julia and Young-Scholten, Martha. Cambridge: Cambridge University Press, 372-393.

Russell, Jane and Spada, Nina. 2006. The Effectiveness of Corrective Feedback for the Acquisition of L2 Grammar. A Meta-Analysis of the Research. In: *Synthesizing Research on Language Learning and Teaching*, ed. by Norris, John M. and Ortega, Lourdes. Amsterdam: John Benjamins, 133-164.

Salaberry, M. Rafael. 2000. The Acquisition of English Past Tense in an Instructional Setting. *System*, 28 (1), 135-152.

Salkind, Neil J. 2006. *Tests & Measurement for People Who (Think They) Hate Tests & Measurement*. Thousand Oaks: SAGE.

Salkind, Neil J. 2008. *Statistics for People Who (Think They) Hate Statistics*. 3rd ed. Los Angeles: SAGE.

Samad, Arshad Abd and Nurusus, Elisha. 2015. The Influence of Teaching Experience, School Location and Academic Background on Teachers' Beliefs in Teaching Grammar. *Asian Social Science*, 11 (15), 257-266.

Sartoneva, Pirkko (ed.). 1998. *Vieraiden kielten osaaminen Suomessa – aikuisten kielitaidon arvointi.* Arviointi 6/98. Helsinki: Opetushallitus.

Saville-Troike, Muriel. 2012. *Introducing Second Language Acquisition*. 2nd ed. Cambridge: Cambridge University Press.

Schachter, Jacquelyn. 1998. Recent Research in Language Learning Studies: Promises and Problems. *Language Learning* 48 (4), 557-583.

Sewell, Andrew. 2010. Research Methods and Intelligibility Studies. *World Englishes*, 29 (2), 257-269.

Shibata, Miki. 2010. How Japanese Teachers of English Perceive Non-Native Assistant Teachers. *System*, 38 (1), 124-133.

Singleton, David and Muñoz, Carmen. 2011. Around and Beyond the Critical Period Hypothesis. In: *Handbook of Research in Second Language Teaching and Learning: Volume II*, ed. by Hinkel, Eli. New York: Routledge, 407-425.

Sjöholm, Kaj. 1983. Problems in "Measuring" L2-Learning Strategies. In: *Psycholinguistics and Foreign Language Learning: Papers from a Conference Held in Stockholm and Åbo, October 25-26, 1982*, ed. by Ringbom, Håkan. Åbo: Publications of the Research Institute of the Åbo Akademi Foundation, 174-200.

Sorace, Antonella. 2003. Near-Nativeness. In: *The Handbook of Second Language Acquisition*, ed. by Doughty, Catherine J. and Long, Michael H. Malden: Blackwell, 130-151.

Statistics Finland. 2016. *Population According to Language 1980-2015*. Available at: www.tilastokeskus.fi/til/vaerak/2015/vaerak_2015_2016-04-01_tau_002_en.html (accessed on 29 November 2016).

Storch, Neomy. 2015. Researching Grammar. In: *Research Methods in Applied Linguistics: A Practical Resource*, ed. by Paltridge, Brian and Phakiti, Aek. London: Bloomsbury, 349-368.

Swan, Michael. 2002. Seven Bad Reasons for Teaching Grammar – and Two Good Ones. In: *Methodology in Language Teaching: An Anthology of Current Practice*, ed. by Richards, Jack C. and Renandya, Willy A. Cambridge: Cambridge University Press, 148-152.

Tabatabaei, Omid and Dehghani, Marzieh. 2012. Assessing the Reliability of Grammaticality Judgment Tests. *Procedia – Social and Behavioral Sciences*, 31, 173-182.

Takala, Sauli. 2004. Englannin kielitaidon tasosta Suomessa. In: *Kansainväliset IEA-tutkimukset Suomi-kuvaa luomassa*, ed. by Leimu, Kimmo. Jyväskylä: Koulutuksen tutkimuslaitos, 255-276.

Teleman, Ulf; Hellberg, Staffan and Andersson, Erik. 1999. *Svenska Akademiens grammatik. Vol. 2: Ord, Vol. 3: Fraser and Vol. 4: Satser och meningar.* Stockholm: Svenska Akademien. Also available at: https://svenska.se/SAG.pdf (accessed on 3 July 2018).

Thewissen, Jennifer. 2013. Capturing L2 Accuracy Developmental Patterns: Insights from an Error-Tagged EFL Learner Corpus. *The Modern Language Journal*, 97, 77-101.

Thomas, Margaret. 2013. History of the Study of Second Language Acquisition. In: *The Cambridge Handbook of Second Language Acquisition*, ed. by Herschensohn, Julia and Young-Scholten, Martha. Cambridge: Cambridge University Press, 26-45.

Thornbury, Scott. 1999. *How to Teach Grammar*. Harlow: Longman.

Tremblay, Annie. 2011. Proficiency Assessment Standards in Second Language Acquisition Research: "Clozing" the Gap. *Studies in Second Language Acquisition*, 33 (3), 339-372.

Tuokko, Eeva. 2000. *Peruskoulun 9. vuosiluokan englannin (A1-kieli) oppimistulosten kansallinen arviointi: 1999*. Helsinki: Opetushallitus.

Tuokko, Eeva. 2007. *Mille tasolle perusopetuksen englannin opiskelussa päästään? Perusopetuksen päättövaiheen kansallisen arvioinnin 1999 Eurooppalaisen viitekehyksen taitotasoihin linkitetyt tulokset*. Jyväskylä Studies in Humanities 69. Jyväskylä: Jyväskylän yliopisto.

University of Helsinki Language Centre. 2018. *University of Helsinki Language Centre*. Available at: www.helsinki.fi/en/language-centre (accessed on 25 June 2018).

Ur, Penny. 2011. Grammar Teaching. In: *Handbook of Research in Second Language Teaching and Learning: Volume II*, ed. by Hinkel, Eli. New York: Routledge, 507-522.

Vainikka, Anne and Young-Scholten, Martha. 2013. Stagelike Development and Organic Grammar. In: *The Cambridge Handbook of Second Language Acquisition*, ed. by Herschensohn, Julia and Young-Scholten, Martha. Cambridge: Cambridge University Press, 581-604.

Vogt, W. Paul. 2007. *Quantitative Research Methods for Professionals*. Boston: Pearson.

Wagner, Elvis. 2015. Survey Research. In: *Research Methods in Applied Linguistics: A Practical Resource*, ed. by Paltridge, Brian and Phakiti, Aek. London: Bloomsbury, 83-99.

Watson, Annabel. 2012. Navigating 'the Pit of Doom': Affective Responses to Teaching 'Grammar'. *English in Education*, 46 (1), 22-37.

Watson, Annabel Mary. 2015. Conceptualisatins of 'Grammar Teaching': L1 English Teachers' Beliefs About Teaching Grammar for Writing. *Language Awareness*, 24 (1), 1-14.

Whong, Melinda and Wright, Clare. 2013. Scope and Research Methodologies. In: *The Cambridge Handbook of Second Language Acquisition*, ed. by Herschensohn, Julia and Young-Scholten, Martha. Cambridge: Cambridge University Press, 71-91.

Ylioppilastutkintolautakunta. n.d. *The Finnish Matriculation Examination*. Available at: www.ylioppilastutkinto.fi/en/ (accessed on 7 May 2017).

Yule, George. 2014. *The Study of Language*. 5th ed. Cambridge: Cambridge University Press.

Appendices

Appendix 1. The English-language translation of the questionnaire given to the students. The originals were in Finnish and in Swedish.

Test in the English language

The target group for this test is university-level students who have attended comprehensive school and upper secondary school in Finland and studied English from primary school (as a so-called A1 or A2 language).

The results will be used for as the corpus for a doctoral dissertation. You should not indicate your name when you answer the questionnaire. There are three parts in the test: background information, the actual test and self-assessment.

While answering, please do not use any aids (such as grammar books or dictionaries) and do not discuss the answers with others. Do not spend too long in considering the answers and do not change your answers when working on the self-assessment part. Please hand in the papers to your own teacher or leave them in my pigeonhole in the Language Centre of the University of Helsinki, third floor.

If necessary, feel free to continue writing overleaf.

THANK YOU VERY MUCH FOR YOUR PARTICIPATION!

Satu Pesola
English language teacher
University of Helsinki Language Centre

In case you are willing to volunteer to take part in an additional study involving either a writing task, an interview or both, please give your contact details below and I will contact you. Thank you already beforehand!

Name:
_____

E-mail address:
_____

Phone number:
_____

Background information for the test in English

Gender male female

First language Finnish Swedish Another language, which? _____

Year of birth _____

Year of finishing upper secondary school _____ spring autumn

In which grade did you start studying English at school? _____

How many years have you studied at a university? _____

Main subject _____

Minor subject(s) _____

Mark in English in the matriculation examination and the school-leaving certificate _____

Mark in the L1 in the matriculation examination and the school-leaving certificate _____

Which compulsory courses or examinations in English have you taken at the university? ____

_____

Additional / voluntary studies in English (what, where, when, for how long) _____

_____

_____

_____

_____

_____

Visits to English-speaking countries for longer than a month (what, where, when, for how long)_____

_____

_____

_____

Skills in other languages (how many years, when and where have you studied them) _____

_____

_____

_____

Any other information that might influence your test results _____

_____

_____

Self-assessment part

1. How much time did you need to complete the test? _____

2. What is your own opinion on how well you did? What do you think you did well, what was difficult? _____

_____

_____

_____

_____

_____

_____

_____

_____

_____

3. What criteria do you use when forming and choosing verb forms in English? _____

_____

_____

_____

_____

_____

_____

_____

_____

4. What kind of memories do you have from lessons focusing on English grammar at school?

_____

_____

_____

_____

_____

_____

_____

_____

_____

Murphy & Hashemi

English Grammar in Use Supplementary Exercises

Verb forms: revision

Can you put the verbs in the correct form and solve this detective puzzle?

TREVOR STERN was not a popular man, in spite of his wealth. He (1) lived (live) in a large house about a mile outside the village of Prenton. When he (2) was found (find) dead in his study, no one (3) cried (cry), not even his only daughter. It was soon clear that he (4) ..... (murder).

Detective Inspector Blackledge took statements from his widow, Dorothy, his seventeen-year-old daughter, Lucy, his business partner, Gerald Brook, and his doctor.

I (5) ..... (not/love) my husband, he was a cold and selfish man. But I (6) ..... (not/murder) him, either. After dinner last night he said he (7) ..... (want) to check some business papers in his study. He (8) ..... (have) a meeting with Gerald, his business partner, the next morning. He (9) ..... (ask) for some tea. That was about 9 o'clock. I (10) ..... (watch) a rather exciting film on television, so I (11) ..... (tell) Lucy to take it to him. At quarter past nine Doctor Emerson (12) ..... (call). I (13) ..... (notice) the time because we (14) ..... (expect) him to come earlier. I (15) ..... (answer) the front door bell. Trevor (16) ..... (still/shout) in his study. He and Lucy (17) ..... (obviously/have) a serious row. So I (18) ..... (take) the doctor into the sitting-room for a moment. Then Trevor stopped (19) ..... (shout). I guessed Lucy (20) ..... (go) out by the back door. Doctor Emerson went to the study. I think he wanted to persuade Trevor (21) ..... (go) to the hospital for some tests, but Trevor (22) ..... (not/want) to go. I (23) ..... (hear) him shouting again several times over the next twenty minutes. He called him an ignorant country doctor, and later he said something like 'There's nothing you can do!' I think Lucy (24) ..... (come) into the house while the doctor (25) ..... (still/talk) to Trevor. I (26) ..... (hear) the front door bang during a quiet few seconds when Trevor (27) ..... (not/shout). I was tired and fed up and went to my bedroom soon after that. My sister (28) ..... (phone) and we (29) ..... (talk) for ages. I (30) ..... (tell) her I (31) ..... (decide) to leave Trevor.

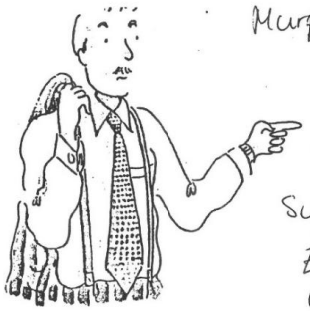
Appendix 2. The test given to the students. Written permission to use the test was obtained from Cambridge University Press on 11 August 2003.

Murphy & Hashemi: English Grammar in Use Supplementary Exercises

Mum (32) ..... (watch) some stupid film after dinner, so she made me (33) ..... (take) Dad's tea into his study. It was about nine o'clock. He was in a really mean mood. He shouted at me because I (34) ..... (spill) a few drops of tea on his desk while I (35) ..... (pour) it. I (36) ..... (not/want) to watch the film so I (37) ..... (creep) out by the back door. I (38) ..... (decide) to go down to the village and use the public phone to call Alan. He's my boyfriend. I (39) ..... (never/like) Mum or Dad to be around when I (40) ..... (talk) to him. Especially yesterday, because Dad and I (41) ..... (have) a stupid argument about Alan the day before. It (42) ..... (normally/take) quarter of an hour to walk to the village. Perhaps it (43) ..... (take) less time last night. I can't prove I (44) ..... (go) to the village. No one (45) ..... (see) me when I (46) ..... (walk) into the village. I (47) ..... (see) Gerald, that's Dad's business partner. He (48) ..... (stand) near the window in his sitting-room. He (49) ..... (not/see) me, though, because it was dark outside. He (50) ..... (talk) on the phone, I think.

Alan (51) ..... (not/answer) the phone. Then I (52) ..... (remember) he (53) ..... (tell) me he (54) ..... (play) in a concert that evening. So I (55) ..... (walk) home again. I (56) ..... (meet) Gerald just before I (57) ..... (reach) our house. He (58) ..... (look) for his dog. That was about twenty to ten. I came in by the back door as quietly as possible and went to bed. I didn't want to see my parents again that evening.

I (59) ..... (call) at the Sterns' house at nine-fifteen. I (60) ..... (be) rather later than I (61) ..... (plan) to be because I (62) ..... (visit) another patient. When Mrs Stern (63) ..... (let) me into the house she (64) ..... (seem) rather embarrassed and (65) ..... (show) me into the sitting-room. I could hear Trevor Stern (66) ..... (shout) at someone in his study. Mrs Stern said something about teenage girls and that they (67) ..... (have) problems with Lucy. Well, the shouting (68) ..... (stop) almost immediately, so I (69) ..... (go) to his study. Lucy (70) ..... (already/leave) the room before I (71) ..... (get) there. I tried (72) ..... (explain) to Trevor why he needed (73) ..... (have) these hospital tests, but he (74) ..... (not/let) me. He said I (75) ..... (be) an ignorant country doctor who (76) ..... (not/know) what he (77) ..... (talk) about. I (78) ..... (realise) it was no use (79) ..... (argue) with him so I (80) ..... (leave) after only a few minutes. I was quite angry actually. I let myself out of the house without (81) ..... (see) Lucy or Mrs Stern.



Yes, Trevor was my business partner. We (82) ..... (not/be) really friends. Yes, my house (83) ..... (be) just round the corner from the Sterns'. I (84) ..... (live) here for two years now. I (85) ..... (have) a little cottage in the village. But I (86) ..... (buy) this house when I started (87) ..... (earn) a lot of money.

I can't really tell you very much about the night Trevor died. I took my dog for a long walk that evening. I (88) ..... (go) up on the hills, away from the village. Then the stupid dog (89) ..... (go) after a rabbit or something and I (90) ..... (lose) him in the dark. I (91) ..... (look) for him when I (92) ..... (meet) Lucy, as a matter of fact. She (93) ..... (walk) up the road towards their house. She (94) ..... (seem) rather upset. I asked her if she (95) ..... (see) the dog, but she said she (96) ..... (not/had). She (97) ..... (go) into her house and I (98) ..... (find) him a few minutes afterwards. I was back home by just after quarter to ten.

Detective Inspector Blackledge showed the statements to her colleague, Sergeant Ross.

BLACKLEDGE: Well, Ross. What do you think? Who killed Stern?

ROSS: I don't know. It (99) ..... (not/be) his wife. She (100) ..... (not/even/go) into the study.

BLACKLEDGE: But she admits she didn't love him. Do you think she's in love with the doctor?

ROSS: It's possible. And perhaps Trevor Stern (101) ..... (find out). But we know the doctor was at the hospital by ten o'clock that night. And that's at least half an hour from the Sterns' house.

BLACKLEDGE: But that (102) ..... (mean) he (103) ..... (leave) the Sterns' house before half past nine.

ROSS: Exactly.

BLACKLEDGE: Anyway, Dorothy Stern told her sister she (104) ..... (leave) her husband. She didn't need (105) ..... (murder) him.

ROSS: But what about Lucy?

BLACKLEDGE: Yes, there's something about Lucy's story which doesn't quite fit. Let's see, what did Gerald Brook say?

ROSS: That's it! Lucy (106) ..... (not/walk) to the village and back, if he (107) ..... (meet) her at twenty to ten. She (108) ..... (still/shout at) by her father at nine-fifteen.

BLACKLEDGE: But look at all the statements. The times don't fit.

ROSS: Neither do the facts. Someone (109) ..... (tell) lies.

BLACKLEDGE: I think it's time we (110) ..... (make) an arrest.

Appendix 3. The recruitment and instructions sheet given to the teachers.

Dear colleague, Helsinki, September 8th, 2004

I am pursuing a doctor's degree in English at the Department of English at the University of Helsinki under the supervision of Professor Terttu Nevalainen and PhD Minna Palander-Collin. The purpose of my research is to find out how well Finnish university students master the use of English verb forms and what kind of variation native or native-like speakers of English allow in the use of these forms.

I am now asking for your help in order to find out the range of variation allowed. I would be grateful if you had the time to fill in the attached questionnaire. I know this is asking you a lot, but I would really appreciate your help.

Here are some instructions:
Enclosed please find three sections:
a) "verb forms: revision",
b) "alternatives given for the test" and
c) "background information".
The students were given the test "verb forms: revision" and they were asked to fill in the correct forms. I have collected all the answers that these students (ca. 350 in total) gave in each slot and listed them in the "alternatives given for the test" in alphabetical order. I would now like to ask you to rank these answers using the following scale:

1 - this is the best alternative (if necessary, you can give several 'best' answers)
2 - this is an acceptable alternative, but not the most suitable
3 - this is a questionable alternative (you are not sure if one can use this form or not)
4 - this is an inappropriate alternative

Please compare the students' answers with the test text and decide how you would evaluate each alternative. Please indicate your choice after each alternative in each slot. After that, please fill in the "background information" sheet.

An example:

9.

asked	1	askt	4	has asked	4	was asking	3
asked me	2	did ask	3	is going to ask	4		
asks	4	had asked	2	was asked	4		

Please return your contribution via internal mail to the address below. I would be grateful if you could do this by October 15th, 2004. If you have any questions, please contact me at the address or number below.

I really appreciate your help. Thank you so much!

Best regards,

Satu von Boehm
Language Centre
University of Helsinki
P.O. Box 4 (Fabianinkatu 26)
00014 University of Helsinki
satu.vonboehm@helsinki.fi
tel. 09 191 23 187

Appendix 4. The complete list of verb forms supplied by the students given to the teachers.

Alternatives given for the test

4.
had become
murdered
had been murded
had been murder
had been
murdered
had been
murderer
had murded
had murdered
has been
murdered
have been
murdered
is a murder
it was a murder
murdered
not murdered
was been
murdered
was murded
was murder
was murderd
was murdered
was murdured
was murdered by
his family
member
was the murderer
was to murder

5.
am not in love
with
did not
did not love
did not loved
didn't
didn't love
didn't loved
do not love
don't love
dont
have not loved
I did not love
I did not loved
I'm not in love
with
not in love with
was not in love
was not love
was not loving
wasn't in love
wasn't in love
with

6.
did not
did not murded
did not murder
did not murdered
did'n murder
didn't kill
didn't murd
didn't murded
didn't murder
didn't murdered
didn't murdet
didn't slaughter
don't murder
have not been
murdered
have not
murdered
wasn't murder

7.
that he wanted
want
wanted
wanted to
wantit
wants
wated
wished
would want

8.
did have
had
had had
has
have
have had
is going to have
is having
should have
was about to have
was going to
have
was having
was supposed to
have
was to have
will have
would had have
would have
would have had
would've had

9.
asked
asked me
asks
askt
did ask
had asked
has asked
is going to ask
was asked
was asking

10.
did watch
had watched
wacthed
waled
was waching
was watching
watced
watch
watched
watching

11.
tell
telled
told
told to
toll

12.
call
called
called me
calls
have called
made a call

13.
did not notice
did notice
didn't notice
had noticed
have noticed
notice
noticed
notised

14.
are expected
did not expect
didn't expect
expect
expected
expecting
had been
expecting
had expected
have been
expecting
have expected
have expecting
was expected
was expecting
were expected
were expecting
where expecting
where extecting

15.
answed
answeder

answer
answerd
answered
anwered
astered
aswered
did answer
was answered

16.
had been still shouting
have still shout
is still shouting
shout still
shouted
shouted still
still beeing shouting
still didn't shout
still shout
still shouted
still shouthed
still shouting
still shouts
still was shouting
was shouting
was shouting still
was still shoting
was still shounting
was still shout
was still shouted
was still shoutig
was still shouting
was still shoutting
was still shut
were still shouting
was still and shouted

17.
are obviously having
bad obviously
did obviously
have
had
had obivously
had obviosly
had obviouslu
had obviously
had obviously
had
had obviously have
has obviosly
has obviously have
have obviously
have obviously
had
obiosly had
obiously had
obviosly did have
obviosly had
obviosly had had
obviosly has
obvioulsly had had
obviously ad
obviously ave
obviously did have
obviously had
obviously had
had
obviously have
obviously have
had
obviously was having
obviously were having
obviousy had
oviously had
seemed to have
was obviously having
were obiously having
were obviously having
were obviouslyt having

18.
did take
had been taken
take
taked
toke
toked
took
tooked
touk
tuok

19.
and shouted
for shouting
shout
shouted
shouthing
shouring
shouting
shoutting
the shouting
to shout
to shouth

20.
did go
get
gick
go
goes
going
gone
got
had gone
had gonne
had got
had went
has go
has got
have went
to go
to went
was going
was gone
went
will go
would go

21.
go
goes
going
had went
into going
should go
to get
to go
to go to
was going
went
would go

22.
did not
did not want
did not wanted
didn't wait
didn't want
didn't wanted
didn't wont
didnt want
do not want
do not wanted
doesn't want
don't want
not wanted
wanted not
will not wont

23.
did hear
did heard
had heard
have heard
hear
heard
heared
heart

24.
came
comed
did come
had came
had come
had comen
has been come
have come
to come
was coming
went

25.
are still talking
had been still talking
still talk
still talked
still talking

still talks
still was talking
still were talking
talked still
was still talking
was talking
were still talking
were talking
where still talking

26.
did hear
did heard
did'n hear
h
had heard
had heared
hear
heard
hearded
heared
hearg
heart
was hearing

27.
did not shoud
did not shout
did not shoutt
didn't shout
didn't shouted
didn't shut
doesn't shout
don't shout
had not shouted
hadn't shout
stoped shouting
stopped shouting
was not shouted
was not
shouthing
was not shoutidg
was not shouting
was not shoutting
was'nt shuuting
wasn't shout
wasn't shooting
wasn't shouting
wasnt shouting
weren't shouting
were not shouting

28.
call
called
called me
had phoned
phone
phoned
phoned me
phoned to me
rang
ringed

29.
did talking
had talked
hadn't talked
talk
talked
tolk
was talking
were talked
were talking
were talkingk

30.
did tell
had told
tald
tell
telled
tod
todd
told
was telling

31.
've decided
decide
decided
decidid
hac decided
had been decided
had decide
had decided
had decides
had decited
hade decided
has been decided
has decided
have decided
have decidet
have decited
have dicided
to decide
was decide
was decided
was deciding

32.
did watch
had watch
had watched
is watching
wached
was watchig
was watching
watced
watched
were watching

33.
has taken
take
taked
taking
to take
to took
took

34.
bad spilled
did spill
had been spilled
had speld
had spill
had spilled
har spilled
have spilled
have spoilt
I had spilled
spell
spill
spilld
spilled
spillt
spilt
spoil
was spilled
was spilling

35.
bring
did pour
did pouring
had been pourid
had poured
pour
pour of
pourd
pourded
poured
poured it
pouring
pourred
pourted
was poured
was pouring
was pourring
were pouring

36.
didn't
did not want
did not wanted
did not wont
did'n want
did'n't wanted
didn it want
didn't want
didn't want to
didn't want to
watch
didn't wanted
didn'twant
didnt want
don't want
won't wont

37.
cept
chept
creaped
creapt
creep
creepd
creeped
creept
crep
creped
crept
crooped
crop
was creeped

38.
decicided

decide
decided
dicided
had decided
had decited
have decided

39.
did never like
didn't like
do never like
don't ever like
don't like
dont never like
had never liked
have never likd
have never like
have never liked
haven't ever liked
like never
never did like
never did liked
never have liked
never like
never like do
never like for
never liked
was never liked

40.
'm talking
am talk
am talking
an talking
did talk
had talked
have talked
have talken
speak
spoke
taking
talk
talke
talked
talking
tolk
took
was talked
was talking

41.
did have
had
had had
had have
has
have
have had
have hah
were having

42.
do normally
taken
had normally
taken
has normally
taken
normally take
normally taked
normally taken
normally takes
normally takes a
normally toked
normally took
normally tooked
normally touk
normally was
taking
normaly takes
take normally
takes normally
takes normally a
took normally
was normally
took
will normally
take
will take
normally
would normally
take

43.
did take
had been taken
had tooked
take
taked
taken
takes
toke
took
took me
tooked
touk
tuke
tuok
was taken
was taking
were taking
would take

44.
am going
did not go
get
gick
go
going
gone
got
had gone
had went
hade gone
have gone
was going
was gone
wen't
went
will go
wouldn't have
gone

45.
couldn't
couldn't see
didn't
didn't see
did saw
had seen
hadn't seen
have seen
saw
sawed
saws
see
seen
sees
sow
was seeing

46.
did walking
walk
walked
walking
was walking
were walking

47.
did saw
did see
saw
sawed
sed
see
sow

48.
stand
standed
stands
stod
stood
stoud
was sitting
was standing
were standing

49.
cant see
couldn't see
didn't
did not saw
did not see
didn't sa
didn't saw
didn't see
didn't seen
didn'tsee
doesn't see
dosen't saw
had not seen
hadn't seen
not saw

50.
talked
talks
tolked
was talkin
was talking

51.
answers
did not

did not answed
did not answer
did not answered
did not aswer
didn't
didn't answer
didn't answerd
didn't answered
didn't aswered
doesn't answered
dosen't answered
hadn't answered
havent nt
answered
wasn't answering
was not
answering

52.
did remember
do remember
remember
rememberd
remembered
remembered that
remembert that
remembred
remenbered

53.
tells
had told
had told to
had tolld
has told
talk
telling
to tell
told
told me
was telling
was told

54.
'd play
'll play
are playing
be playing
going to play
had been plaing
had been playing
had played
had to play
is going to play
is playing
plaied
play
playd
played
playes
plays
should play
to play
was going to play
was paying
was plaing
was playing
was to play
were playing
will be playing
will play
woold play
woud play
would be playing
would had been
playing
would play

55.
had to walk
waked
walk
walked
went

56.
had met
meat
meet
meeted
met
mett

57.
had reached
reach
reached
reachhed
reaching
reacht
was reach
was reaching

58.
are looking
is looking
looced
look
looked
looks
was looking
was looking for
was lookking
was lookng
was woking
were looking

59.
call
called
did call
had a call
had called
was called
was calling

60.
'd be
'll be
am
be
been
did be
had been
have
have been
was
was beeing
was being
was going to be
was going to be
there
was there
were
were there
will be
would be
would have been

61.
'd planed
did plan
did planing
had plaint
had plan
had planed
had planned
hade planned
have planed
have planned
plan
planed
plannd
planned
planted
was planed
was planing
was planned
was planning

62.
had been visiting
had had to visit
had to visit
had visit
had visited
viseted
visit
visited
was going to visit
was visited
was visiting
were visiting
would have
visited

63.
did let
had let
led
left
let
lets
was letting

64.
might seem
saw
seem
seemd
seemed
seemed to be
seems
seems to be
seemt
was seeming

65.

shew
show
showd
showed
showm
shown
was showing

66.
had been shouted
shouiting
shout
shouted
shouting
shoutting
to shout
was
was shouted
was shouting

67.
are having
did have
had
had had
had have
had some
has
have
have been
have had
having
was having
were having
where having

68.
had stopped
stop
stoped
stopid
stopped
stopt
was stoped

69.
did go
get
gick
go
goes
going
gone
got
was going
went

70.
already had
leeved
already had left
already leaved
already leaves
already leaving
already left
had allready
leaved
had allready left
had alreade
leaved
had alreade left
had already leave
had already
leaved
had already left
had alredy leaved
had alredy left
has already leave
has already
leaved
has already
leaven
has already left
has been already
leave
have already
leaved
have already left
have left
leaved already
left already
was allready left
was already leave
was already
leaved
was already
leaving
was already left
was already lived

71.
get
got
got in
gott
had gotten
was getting
went
will got

72.
explain
explained
explaining
to exlpain
to explain
to explaind

73.
had
had have
has
have
have to have
having
to have

74.
did not let
did not let me
didn’t
didn’t let
didn’t let me
didnt let
doesen’t let
doesn’t let
don’t let
not let
won’t let
would not let
would not let me
wouldn’t let

75.
‘am
am
be
had been
has been
have
have been
m
to be
war
was
was beeing
was being
were
will be
would be

76.
did not know
did not known
didn’t
didn’t knew
didn’t know
didnt
do not know
doen’t know
does
does not know
does’n’t know
doesn’t
doesn’t know
don’t know
was not knowing

77.
had talked
is talking
is talkings
talk
talked
talked about
talking
talking about
talks
was
was talked
was talking
was talking about
were talking

78.
had realised
realise
realised
realised that
realized
relised

79.
argue
argued
argueeing
argueing
arguing

argureering
of arguing
to argue
to argued
to arque

80.
did leave
leaft
leaved
left
levt
was leaving

81.
being seen
saw
sawing
see
seeing
seing
to see

82.
are not
aren't
did not be
did'nt be
didn't be
didn't was
didn't were
had not been
hadn't been
have been not
haven't be
not are
wasen't
wasn't
wasn't been
were no been
were not
were not a
were not being
weren:t
weren't
werent
wern't
where not

83.
being
is
ist
lies
was
were

84.
've lived
am living
had been living
had lived
has live
has lived
have been lived
have been living
have lied
have live
have lived
have living
lived

85.
've got
also have
do have
had
had have
has
havd
have
have had
used to have

86.
bought
buy
buyd
buyed
have bought
will buy

87.
earn
earned
earnig
earning
to ean
to earn

88.
did go
get
gick
go
goes
gone
got
had gone
have went
used to go
was going
wen't
went

89.
did go
get
gick
go
goes
got
had gone
left
was going
wen't
went
wents

90.
did lose
loose
loosed
lose
losed
lost
losted
lust

91.
have been
looking
looked
took a look
was locking
was looking
were looking

92.
meat
meet
meeted
met
mete
mett
saw
was meeting

93.
had walked
walked
walked up
walking
was waking
was walking

94.
did seem
seamed
seem
seemd
seemed
seemed to be
seems
seems to be
seemt
sem
showed to be
was seeming

95.
had been
had been seeing
had saw
had sawn
had seen
has been saw
has saw
has seen
have been seen
have saw
have seen
save
saw
saws
see
seen
sees
was saw
was seen
would had seen
would have seen

96.
did not
did not had
did not have
didn't

didn't had
didn't had seen
didn't has
didn't have
didnt have
din't had
had not
had not had
had not seen
had not seen it
hadn't
hadn't had
hadn't have
hadn't seen
hadn't seen it
hadnt
has not
hasen't
hasn't
hasn't had
hasn't have
have not had
haven't
haven't had
would not had

97.
did go
get
gick
goes
gone
got
had gone
was going
wen
wennt
went

98.
did find
find
finded
foud
found
founded
foundt
fount
was finding

99.
am not
can be
can not be
can not have been
can't be
cannot be
could not
could not be
could not have been
couldn't be
couldn't have been
didn't be
didn't was
does not be
had not been
hadn't been
has not been
has to be
hasn't been
is not
isn't
isnt
was not
was'nt
wasn't
wasn't be
wasn't been
wont be
wouldn't be

100.
did not even
did not even get
did not even go
did not even to go
did not even went
did not go even
did'n even go
didn't even
didn't even go
didn't even gone
didn't even went
didn't even went to
didn't ever go
didn't evnt go
didnt even go
didntn even went
dindn't even go
dinn't even go
ditn't even go
does not even go
doesen't even get
doesn't even go
don't go ween
dosent even go
even did not go
even didn't go
even didn't went
even doesn't went
even gone
even went
had not even go
had not even gone
had not even got
hadn't even go
hadn't even gone
hadn't even went
has not even gone
has not even went
hasn't
hasn't even go
haven't even went
haven't event gone
never went
not even went
was even not
wasn't even going
won't even go

101.
did find it out
did find out
find out
finded out
finds out
fond out
fonded out
fouded out
found it out
found out
found out it
found that out
founded out
had found
had found it out
had found out
had found that out
had founded out
has find out
have finded it out
have found it out
was find out that
was fouding
was found out
will be found out
will find out
would find out

102.
could mean
didn't mean
does mean
doesn't
doesn't mean
is meaning
mead
mean
meane
meaned
meaning
means
means that
meant
menang
ment
met
will mean
would mean
would meen

103.
did leave
didn't leave
had le
had leaved
had left
had to leave
has leave
has left
have left
leave
leaved
leaven
leaves
left
levt
must have left

should have left
would have left

104.
'd leave
'll leave
are going to leave
can't leave
didn't leave
going to leave
gonna leave
had leaved
had leaving
had left
had levt
has left
is going to leave
is going to
leaving
is leaving
leave
leaved
leaves
left
wants to leave
was going to
was going to
leave
was leaving
was left
was planning to
leave
was to leave
were leaving
will leave
would be leaving
would leave

105.
murder
murdered
murdering
to have murdered
to murde
to murded
to murden
to murder

106.
can't have
walked
cannot have
walked
could not have
walked
could not walk
couldn't have
walked
couln't walk
did not
did not walk
didn't
didn't had walk
didn't walk
didn't walked
didnt walk
don't walk
dosen't walk
had not walked
hadn't walk
hadn't walked
has not walked
hasn't walked
not walk
not walked
walked not
was not walking
wasn't
wasn't walk
wasn't walked
wasn't walking
weren't walking
would not have
walked
wouldn't have
walked
wouldn't walk

107.
didn't meet
had met
had mett
has met
meat
meet
meets
met
saw
was going to
was meeting

108.
had been
shouting at
had still been
shouted at
had still shout at
has still been
shouted at
have still shouted
is still shout at
is still shouted at
should be shouted
at
shouted at still
shouted still at
shouted stll at
still did shout at
still shoted at
still shout
still shout at
still shouted
still shouted at
still shouted by
still shouts at
still was shouting
at
was being
shouted at
was still at
shouted
was still beeing
shouted at
was still being
shout at
was still being
shouted at
was still being
shouting at
was still getting
shouted at
was still shout at
was still shouted
was still shouted
at
was still shouted
up
was still shouthed
was still shoutig
at
was still shoutin
at
was still shouting
was still shouting
at
was still
shoutting at
was till shouting
was till shouting
at
were still shouted
at
were still
shouting at
would have been
still shouting at
would still be
shouted at

109.
are telling
did tell
had told
has been telling
has told
has told us
have told
is lying
is teling
is telling
is telling lies
tell
tells
told
told us
was telling

110.
'll make
are going to make
are making
can make
do
have made
have to made
have to make
made
make
maked
making
should make
to make
will be making
will make

Appendix 5. The background information sheet given to teacher participants

Background information

Your gender: male female

Your native language: English Finnish Swedish Other: _____

What is/was the native language of your parents? _____

_____

What is the native language of your children (if you have any)? _____

_____

What is the native language of your husband/wife/partner (if you have one)? _____

_____

Your year of birth: _____

In what country and city were you born and where did you grow up? _____

_____

_____

Graduation years: BA or equivalent in _____, MA or equivalent in _____

Name and place of university _____

_____

Major subject at university _____

Minor subject(s) at university _____

_____

Do you have a doctorate or licentiate's degree? If yes, please specify (when, where, what field and topic). _____

_____

_____

What is your profession? _____

_____

How did you learn English in the first place and how do you maintain your skills? _____

_____

_____

Have you got experience in teaching English at university level? Please specify (how long, where, what) _____

_____

_____

Have you got experience in teaching English at other levels? Please specify. _____

_____

_____

Have you taught English grammar? If yes, please specify (to whom, where, when). _____

_____

_____

_____

What other languages do you know, how well can you use them and how did you learn them? How often do you use them? _____

_____

_____

_____

_____

_____

How long have you lived in Finland? _____

_____

_____

How many years in total have you lived in English-speaking countries? _____

_____

_____

_____

When running errands in Finland, what language do you use? _____

_____

How would you describe your attitudes towards learning, teaching and applying English grammar? _____

_____

_____

_____

_____

Any other information that might be important or useful _____

_____

_____

_____

Appendix 6. Alternatives given for the test, with the number of students providing the form and the teachers' assessment

1 – the best, 2 – acceptable, 3 – questionable, 4 – inappropriate

Slot 4

had become murdered	1	3232434444444
had been murded	1	4444444441444
had been murder	2	4444444444444
had been murdered	68	1111111114111
had been murderer	1	4444444444444
had murded	1	4444444444444
had murdered	3	4444444444434
has been murdered	4	44?4444444444
have been murdered	1	4444444444444
is a murder	-	4444444444444
it was a murder	1	4444444444444
murdered	4	4444444444434
not murdered	-	4444444444444
was been murdered	1	4444444444444
was murded	7	4444444444444
was murder	1	4444444444444
was murderd	1	4444444444424
was murdered	173	2424234224422
was murdured	1	44444?4444424
was murdered by his family member	-	44344344233??
was the murderer	2	4444444444434
was to murder	1	4444444444?44
(no answer)	44	-
was murdered (had been murdered)	1	-

Slot 5

am not in love with	1	4444444444422
did not	2	4444444444444
did not love	77	1111111111121
did not loved	-	4444444444444
didn't	3	4444444444444
didn't love	200	1111111111111
didn't loved	2	4444444444??4
do not love	3	4444434444412
don't love	20	4444434444412
dont	-	4444444444?4
have not loved	1	44342344443?2
I did not love	1	4421441122124
I did not loved	-	4444444444444
I'm not in love with	-	4444444444424
not in love with	1	4444444444444
was not in love	-	4432444244444
was not love	1	4444444444444
was not loving	1	4444444444444
wasn't in love	1	4434444444434
wasn't in love with	1	2122244412422
did'nt love	2	-
I didn't love	1	-
'm not in love with / don't love	1	-

Slot 6

did not	1	44?1444444444
did not murded	1	4444444444444
did not murder	78	1111111111111
did not murdered	3	4444444444434
did'n murder	1	4444444444444
didn't kill	1	4231444432424
didn't murd	1	444?444444444
didn't murded	1	4444444444444
didn't murder	204	1111111111131
didn't murdered	17	4444444444444
didn't murdet	-	4444444444444
didn't slaughter	1	4231444434434
don't murder	2	4444444444444
have not been murdered	-	4444444444434
have not murdered	3	4344214224422
wasn't murder	-	4444444444444
did'nt murder	2	-
not murdered	1	-
didn't murder / kill	1	-
I didn't murder	1	-

Slot 7

that he wanted	-	4424414444444
want	18	4444444444444
wanted	265	1111111411111
wanted to	1	4434421422434
wantit	-	4444444444444
wants	31	4144424444444
wated	1	4444444444444
wished	1	4231414434434
would want	2	4224314444444

Slot 8

did have	1	4444444444444
had	219	4131443111111
had had	2	4444444444424
has	10	4444414444434
have	1	4444444444444
have had	1	4444444444444
is going to have	1	4444444444434
is having	1	4434444444434
should have	1	4224344444424
was about to have	1	3122344424422
was going to have	15	1121121111?21
was having	3	4221444411?24
was supposed to have	1	1121112211?21
was to have	4	1111113111?21
will have	6	4444444444?34
would had have	1	4444444444?44
would have	28	4124414444?44
would have had	21	1441224442?31

Slot 8 continues

would've had	1	2441224442?42
had/ would have had	1	-

Slot 9

asked	306	1111111111111
asked me	1	2121212212421
asks	2	4444424444434
askt	-	4444444444444
did ask	2	2134334444442
had asked	2	4224244424222
has asked	-	4444444444444
is going to ask	1	4444444444434
was asked	3	4444444444434
was asking	2	4424334424424

Slot 10

did watch	-	4444344444434
had watched	3	4443324444424
wacthed	1	4444444444444
waled	-	4444444444444
was waching	2	4444444424444
was watching	153	1111111111111
watced	1	4444444444444
watch	5	4444444444434
watched	152	4434344244414
watching	-	4444444444434
walhed	1	-
watchd	1	-

Slot 11

tell	3	4444424444444
telled	2	4444444444444
told	312	1111111111111
told to	1	4444444444444
toll	1	4444444444444

Slot 12

call	3	4444444444444
called	312	1111111111111
called me	-	2221344224424
calls	2	4434424244434
have called	1	4444444444444
made a call	1	4423344434432

Slot 13

did not notice	1	4444444444424
did notice	3	2231224244442
didn't notice	2	4444444444414
had noticed	3	4424444424232
have noticed	1	4434444444434
notice	2	4444424444444
noticed	306	1111111111?1
notised	1	4444444444444

Slot 14

are expected	-	4444444444444
did not expect	1	4444444444434

Slot 14 continues

didn't expect	1	4444444444434
expect	3	4444414444444
expected	182	1114324112211
expecting	1	44?4444444444
had been expecting	3	4121141111221
had expected	44	4111114111111
have been expecting	1	4434444444434
have expected	2	44344?4444444
have expecting	1	4444444444444
was expected	1	4444444444444
was expecting	4	4244444444444
were expected	4	4444441444434
were expecting	65	1124244412421
where expecting	2	4444444444444
where extecting	-	444444?444444
(had) expected	1	-
expeeted	1	-
were excpecting	1	-
were expectig	1	-

Slot 15

answed	1	4444444444444
answeder	-	4444444444444
answer	2	4444424444444
answerd	4	4444444424444
answered	307	1111111111111
anwered	1	4444444444444
astered	-	4444444444444
aswered	2	4444444444444
did answer	-	4244324444444
was answered	1	4444444444444
answeded	1	-

Slot 16

had been still shouting	-	4434344444444
have still shout	1	4444444444444
is still shouting	2	4434424444434
shout still	4	4444444444444
shouted	1	4444444444424
shouted still	13	4234244444444
still beeing shouting	-	4444444444444
still didn't shout	1	4444444444444
still shout	6	4444444444444
still shouted	66	4224444444332
still shouthed	1	4444444444444
still shouting	4	4444444444444
still shouts	-	4444444444444
still was shouting	1	4133444423241
was shouting	3	4134444422234
was shouting still	2	4134434424342
was still shoting	-	4444444444444
was still shounting	1	4444444444444
was still shout	6	4444444444444
was still shouted	2	4444444444444
was still shoutig	-	4444444444444
was still shouting	189	1111111111111

Slot 16 continues

was still shoutting	1	4444444444444
was still shut	1	4444444444444
were still shouting	1	4444444444444
was still and shouted	1	4444444434444
(no answer)	9	-
has still shouting	1	-
was still shouthing	1	-
was till shouting	1	-

Slot 17

are obviously having	-	4444424444434
bad obviously	1	4444444444444
did obviously have	-	3234344444434
had	-	4334444444444
had obivously	-	4444444444444
had obviosly	3	4444444444444
had obviouslu	1	4444444444444
had obviously	56	4134144444432
had obviously had	16	2334142224442
had obviously have	2	4444444444444
has obviosly	1	4444444444444
has obviously	-	4444444444444
have	1	4444444444444
have obviously	2	4444444444444
have obviously had	4	4444444444434
obiosly had	1	4444444444444
obiously had	1	4444444444444
obviosly did have	-	4444344444444
obviosly had	1	4444344444444
obviosly had had	-	4444444424444
obviosly has	1	4444444444444
obvioulsly had had	1	4444444444444
obviously ad	1	4444444444444
obviously ave	1	4444444444444
obviously did have	1	4234344244444
obviously had	139	3234244224432
obviously had had	2	4424244424432
obviously have	3	4444444444444
obviously have had	2	4444444444444
obviously was having	1	4344444444444
obviously were having	1	4123334222141
obviousy had	1	4444444424444
oviously had	1	4444444444444
seemed to have	1	4431444444444
was obviously having	1	4444444424444
were obiously having	1	4444444444444
were obviously having	60	1111111111111
were obviouslyt having	1	4444444424444
(no answer)	2	-
are obviosly having	1	-
had had	1	-

Slot 17 continues

had obviooly	1	-
had obvisouly had	1	-
obiviously had	1	-
obvisiusly were having	1	-
ovsiviosly did have	1	-
were having obviously	1	-

Slot 18

did take	2	4424334444444
had been taken	1	4444444444444
take	3	4444424444444
taked	5	4444444444444
toke	5	4444444444444
toked	1	4444444444444
took	299	1111111111111
tooked	1	4444444444444
touk	2	4444444444444
tuok	-	4444444444444

Slot 19

and shouted	1	4444444444424
for shouting	-	4444444444444
shout	2	4444444444444
shouted	1	4444444444444
shouthing	1	4444444444444
shouring	-	4444444444444
shouting	296	1114111111111
shoutting	3	4444444424444
the shouting	1	4121444223242
to shout	11	4444144444444
to shouth	1	4444444444444
for slouting	1	-
to shout / shouting	1	-

Slot 20

did go	1	4444344444434
get	2	4444444444444
gick	1	444444444?44
go	1	4444444444444
goes	1	4444424444444
going	2	4444444444444
gone	1	4444444444444
got	13	4443444444444
had gone	81	1211111111141
had gonne	-	4444444424234
had got	1	4444444444444
had went	11	4444444444444
has go	1	4444444444444
has got	1	4444444444444
have went	1	4444444444444
to go	31	4444444444444
to went	1	4444444444444
was going	1	2434344224434
was gone	1	4444344444444
went	164	1121234211242
will go	1	4444444444434

Slot 20 continues

would go	1	4444344444444
(no answer)	1	-

Slot 21

go	24	4143444444444
goes	3	4444444444444
going	31	4444444444444
had went	1	4444444444444
into going	6	4121214121434
should go	1	4444344444444
to get	1	4444444444444
to go	239	1111111111111
to go to	1	4414444422224
was going	1	4444444444444
went	7	4444444444444
would go	1	4444344444444
(no answer)	2	-
g	1	-

Slot 22

did not	1	?4?4444444434
did not want	75	1111111111111
did not wanted	1	?444444444444
didn't wait	-	?444444444444
didn't want	223	1111211111432
didn't wanted	5	?444444444444
didn't wont	1	4444444444244
didnt want	2	4444444444244
do not want	1	4444444444424
do not wanted	1	4444444444444
doesn't want	2	4444444444444
don't want	-	4444444444444
not wanted	2	4444444444444
wanted not	1	44444444444442
will not wont	-	4444444444444
didnn't want	1	-
did'nt want	1	-
didn't wan't	1	-
will not want	1	-

Slot 23

did hear	1	2134334444442
did heard	1	4444444444444
had heard	1	4424444444434
have heard	1	4444444444434
hear	3	4434424444444
heard	283	1111111111111
heared	27	4444444444444
heart	1	4444444444444
(no answer)	1	-

Slot 24

came	278	1111211111111
comed	1	4444444444444
did come	1	4234344444444
had came	3	4444444444444
had come	27	4414111114231
had comen	1	4444444444444

Slot 24 continues

has been come	1	4444444444444
have come	-	4444444444444
to come	1	4444444444444
was coming	3	4434444424234
went	1	4444444414424
come	1	-
has come	1	-

Slot 25

are still talking	1	4444444444444
had been still talking	-	4424344444444
still talk	1	4444444444444
still talked	49	4434?44244444
still talking	5	4444444444444
still talks	-	4444444444444
still was talking	4	4143344423232
still were talking	-	4444444444444
talked still	3	4444344444444
was still talking	247	1111111111111
was talking	4	4244444434444
were still talking	3	4444444444444
were talking	1	4444444444444
where still talking	1	4444444444444

Slot 26

did hear	1	4234234444444
did heard	2	4444444444444
did'n hear	-	4444444444444
h	1	444444444?44
had heard	6	1424444124441
had heared	1	4444444444444
hear	3	4444424444444
heard	272	1111111211141
hearded	3	4444444444444
heared	26	4444444444444
hearg	1	4444444444444
heart	1	4444444444444
was hearing	1	4444344444444
I had heard	1	-

Slot 27

did not shoud	1	4444444444444
did not shout	24	2224244444442
did not shoutt	1	4444444444444
didn't shout	115	2222244444422
didn't shouted	3	4444444444444
didn't shut	1	4444444444444
doesn't shout	3	4444444444444
don't shout	1	4444444444444
had not shouted	3	4444444444444
hadn't shout	1	4444444444444
stoped shouting	1	4444444444444
stopped shouting	2	4442342424444
was not shouted	-	4444444444444
was not shouthing	1	4444444421244
was not shoutidg	1	4444444444244
was not shouting	58	1111111114111

Slot 27 continues

was not shoutting	-	4444444424244
was'nt shuuting	-	4444444444444
wasn't shout	1	4444444444444
wasn't shooting	-	4444444444444
wasn't shouting	96	2111111111112
wasnt shouting	1	4444444424244
weren't shouting	1	4444444444444
were not shouting	1	4444444444444
(no answer)	1	-
was'nt shouting	1	-
wasn't shouting / didn't shout	1	-

Slot 28

call	-	44?4444444444
called	5	4121444422421
called me	2	3221444422431
had phoned	3	4424444444444
phone	2	4444444444444
phoned	300	1111111111111
phoned me	4	1221111211421
phoned to me	1	4444444444444
rang	-	4421444422411
ringed	-	4244444444444
phoned / call	1	-
phoned / called	1	-

Slot 29

did talking	1	4444444444444
had talked	1	4444444444444
hadn't talked	1	4444444244424
talk	4	4444424444444
talked	289	1111212111111
tolk	-	4444444444444
was talking	2	4444444444444
were talked	-	4444444444444
were talking	18	4233131114424
were talkingk	-	4444444444444
lalked	1	-
talking	1	-
were talkingc	1	-

Slot 30

did tell	1	4234344444444
had told	1	4424444424444
tald	-	4444444444444
tell	1	4444424444444
telled	1	4444444444444
tod	1	4444444444444
todd	-	4444444444444
told	313	1111111111111
was telling	1	2334344242434

Slot 31

've decided	1	4434424444224
decide	4	4444444444444
decided	78	4424444244424
decidid	-	4444344444444

Slot 31 continues

hac decided	-	4444444444444
had been decided	1	444444444?44
had decide	5	444444444?44
had decided	186	1111111111111
had decides	1	4444444444444
had decited	2	4444444444444
hade decided	1	4444444444444
has been decided	1	4444444444444
has decided	1	4444444444444
have decided	26	4434224444444
have decidet	1	4444444444444
have decited	-	4444444444444
have dicided	2	4444444444444
to decide	-	4444444444444
was decide	1	4444444444444
was decided	6	4444344444444
was deciding	1	4444444444424
(no answer)	1	-

Slot 32

did watch	2	4344344444434
had watch	-	4444444444444
had watched	1	4434344444424
is watching	1	4444434444444
wached	1	4444444444444
was watchig	1	4444444424444
was watching	145	1111111111211
watced	-	4444444444444
watched	164	4233344124114
were watching	3	4444444444444
was watched	1	-

Slot 33

has taken	1	4444444444444
take	172	1111111111111
taked	2	4444444444444
taking	4	4444444444444
to take	100	4144434444444
to took	4	4444444444444
took	31	4444444444444
(no answer)	4	-
totook	1	-

Slot 34

bad spilled	1	4444444444444
did spill	1	4234344444444
had been spilled	1	4444444444444
had speld	1	4444444444444
had spill	2	4444444444444
had spilled	30	4124311114411
har spilled	-	4444444444444
have spilled	-	4444444444444
have spoilt	1	4444444444444
I had spilled	1	4424444424444
spell	2	4444444444444
spill	6	4444444444444
spilld	1	4444444444444
spilled	260	1111111111112

Slot 34 continues

spillt	1	4444444444444
spilt	7	4414141141444
spoil	1	4444444444444
was spilled	-	4444444444444
was spilling	3	4134444424444

Slot 35

bring	-	4444444444444
did pour	-	4444344444444
did pouring	1	4444444444444
had been pourid	1	4444444444444
had poured	1	4444444244444
pour	9	4444444444444
pour of	-	4444444444444
pourd	1	4444444444444
pourded	2	444444444?4
poured	118	1124444424412
poured it	1	4424444244?4
pouring	5	4444444444444
pourred	2	4444444444444
pourted	1	4444444444444
was poured	1	4444444444444
was pouring	170	1111111111111
was pourring	1	4444444444444
were pouring	2	4444444444444
(no answer)	3	-

Slot 36

didn't	1	444?444434244
did not want	65	1111111111111
did not wanted	1	4444444444444
did not wont	1	4444444444444
did'n want	-	4444444424444
did'n't wanted	1	4444444444444
didn it want	1	4444444444444
didn't want	237	2111111111411
didn't want to	1	4414441422414
didn't want to watch	1	4414441422434
didn't wanted	2	4444444444444
didn'twant	-	4444444424444
didnt want	2	4444444424444
don't want	3	4444444444444
won't wont	-	4444444444444
didn't went	1	-
did't want	1	-
won't want	1	-

Slot 37

cept	1	44?4444444444
chept	-	4444444444444
creaped	1	4444444424444
creapt	1	4444444444444
creep	7	4444444444444
creepd	2	4444444424444
creeped	171	4444444414344
creept	2	4444444444444
crep	11	4444444444444

Slot 37 continues

creped	1	4444444444444
crept	117	1111111111111
crooped	-	444444444?44
crop	3	444444444?44
was creeped	1	444444444?44
(no answer)	1	-

Slot 38

decicided	1	4444444444444
decide	2	4444424444444
decided	305	1111111111111
dicided	-	4444444424444
had decided	9	1124211123431
had decited	-	4444444444444
have decided	1	4444444444444
(no answer)	1	-

Slot 39

did never like	2	4144344444444
didn't like	1	4444444444434
do never like	-	4444344424444
don't ever like	1	4434444424434
don't like	1	4424444424434
dont never like	-	4444444444444
had never liked	10	4444441244444
have never likd	1	4444444444444
have never like	3	4444444444444
have never liked	82	4441114121121
haven't ever liked	1	4441444444444
like never	2	4444444441444
never did like	-	4144444411432
never did liked	-	4444444444444
never have liked	2	4442414421231
never like	107	4411224114412
never like do	1	4444444444444
never like for	1	4444444444444
never liked	101	1134412121342
was never liked	1	4444444444444
(no answer)	1	-
(have) never liked	1	-

Slot 40

'm talking	3	4421414411422?
am talk	-	4444444444444
am talking	63	4421214111131
an talking	1	4444444444444
did talk	1	4444444444444
had talked	1	4444444444444
have talked	1	4444444444444
have talken	1	4444444444444
speak	-	4421444424424
spoke	-	4344444444444
taking	1	4444444444444
talk	131	4411114114411
talke	-	44444?4444444
talked	59	1144224444144
talking	4	4444444444444
tolk	1	4444444444444

Slot 40 continues

took	1	4444444444444
was talked	-	4444444444444
was talking	50	1141311421244
am talking / was talking	1	-

Slot 41

did have	2	4434344444424
had	221	1422213214214
had had	58	4111112111121
had have	6	4444444444444
has	2	4444444444444
have	7	4444444444444
have had	19	4444444444434
have hah	-	4444444444444
were having	4	44343444244?4

Slot 42

do normally taken	-	4444444444444
had normally taken	1	4444444444434
has normally taken	1	4144424224434
normally take	6	4444444444444
normally taked	1	4444444444444
normally taken	1	4444444444444
normally takes	207	1111411111314
normally takes a	5	4114111111?41
normally toked	1	4444444444444
normally took	33	2434444214424
normally tooked	-	4444444444444
normally touk	-	4444444444444
normally was taking	1	4444444444444
normaly takes	3	4444444424434
take normally	2	4444444444444
takes normally	35	4434434421444
takes normally a	1	4134234421244
took normally	6	4434444424434
was normally took	1	4444444444444
will normally take	1	4434444244434
will take normally	-	4434444444444
would normally take	8	4121411221332
(no answer)	1	-
noomally takes	1	-
normally took a	1	-
normally will take	1	-
takes normally (a)	1	-

Slot 43

did take	2	4124424444444
had been taken	1	4444444444444
had tooked	1	4444444444444
take	1	4444444444444
taked	6	4444444444444
taken	1	4444444444444
takes	11	4444444444434
toke	3	4444444444444
took	286	1111111111111

Slot 43 continues

took me	1	1311111211334
tooked	2	4444444444444
touk	-	4444444444444
tuke	-	4444444444444
tuok	-	4444444444444
was taken	1	4444444444444
was taking	1	4344444444444
were taking	1	4444444444444
would take	1	4444444444434

Slot 44

am going	2	4444444444444
did not go	1	4444424444444
get	-	4444444444444
gick	1	4444444444444
go	4	4444444444444
going	1	4444444444444
gone	1	4444444444444
got	1	4444444424444
had gone	5	4424344424444
had went	1	4444444444444
hade gone	-	4444444444444
have gone	3	4144444444444
was going	1	4434444224441
was gone	1	4444444444444
wen't	1	4444444444444
went	292	1111111111111
will go	2	4444444444444
wouldn't have gone	1	4444424444434
(no answer)	1	-

Slot 45

couldn't	1	4444444444444
couldn't see	1	4444444444444
didn't	1	4444444444444
didn't see	1	4444444444444
did saw	1	4444444444444
had seen	2	4424444424442
hadn't seen	1	4444444444444
have seen	1	4444444444444
saw	303	1111111111111
sawed	-	4444444444444
saws	1	4444444444444
see	-	4444444444444
seen	1	4444444444444
sees	1	4444424444444
sow	1	4444444444444
was seeing	1	4444444444444
seec	1	-
seem	1	-

Slot 46

did walking	1	4444444444444
walk	3	4444444444444
walked	221	1121114111131
walking	1	4444444444444
was walking	90	2111211111214
were walking	1	4444444444444

Slot 46 continues

walked / was walking	1	-
was walking / walked	1	-

Slot 47

did saw	1	4444444444444
did see	5	2134221241442
saw	308	1111111111?11
sawed	1	4444444444444
sed	-	4444444444444
see	3	4444444444444
sow	1	4444444444444

Slot 48

stand	3	4444444444444
standed	26	4444444444444
stands	1	4444444444444
stod	2	4444444444444
stood	171	4122214444312
stoud	1	4444444444444
was sitting	-	4444444434444
was standing	113	1111111111111
were standing	1	4444444444444
(no answer)	1	-

Slot 49

cant see	-	4444444444444
couldn't see	1	4121111113234
didn't	1	4144444414444
did not saw	2	44444?4444444
did not see	71	1111114111111
didn't sa	1	4444444444444
didn't saw	22	4444444444444
didn't see	214	2111111111111
didn't seen	1	4444444444444
didn'tsee	-	4414444124444
doesn't see	1	4444424444444
dosen't saw	-	4444444444444
had not seen	1	4424444444444
hadn't seen	1	4424444444444
not saw	1	4444444444444
did'n see	1	-
didnt see	1	-

Slot 50

talked	120	4232434444444
talks	1	4444444444444
tolked	-	4444444444444
was talkin	1	4444444424444
was talking	192	1111111111111
(no answer)	1	-
talk	1	-
talked (was talking)	1	-
was talked	1	-
was talkinu	1	-

Slot 51

answers	1	4444444444444
did not	4	4444444444444
did not answed	-	4444444444444
did not answer	65	1111111111111
did not answered	1	44444?4444444
did not aswer	2	4444444444444
didn't	3	4444444444444
didn't answer	223	2111111111111
didn't answerd	1	4444444444444
didn't answered	9	4444444444444
didn't aswered	2	4444444444444
doesn't answered	1	4444444444444
dosen't answered	-	4444444444444
hadn't answered	1	2444444444444
havent nt answered	-	4444444444444
wasn't answering	1	4434444224432
was not answering	1	4434444224432
didnt answer	1	-
didn't answere	1	-
didn't anwer	1	-
haventnt answered	1	-

Slot 52

did remember	2	4434224444444
do remember	-	44444444444?4
remember	17	4444424444444
rememberd	4	4444444424444
remembered	282	1111111111111
remembered that	2	1111111111121
remembert that	1	4444444444444
remembred	1	4444444424444
remenbered	7	4444444444444
remeembered	1	-
remem bered	1	-
rememembered	1	-

Slot 53

tells	1	4444444444444
had told	137	1111111111131
had told to	1	4444444444444
had tolld	-	4444444444444
has told	6	4444444444444
talk	1	4444444444444
telling	5	4444444444444
to tell	1	4444444444444
told	163	4122212111414
told me	1	4422442421444
was telling	1	4434444444444
was told	1	4444444444444
had toud	1	-

Slot 54

'd play	1	4424144444442
'll play	1	4444444444444
are playing	1	4444444444444
be playing	1	4444444444444
going to play	1	4444444444444
had been plaing	1	4444444444444

Slot 54 continues

had been playing	2	4444444444444
had played	1	4444444444424
had to play	2	4421224224144
is going to play	1	4134424444444
is playing	11	4134424444444
plaied	1	4444444444444
play	4	4444444444444
playd	1	4444444444444
played	47	1441444444444
playes	2	4444444444444
plays	17	4144424444434
should play	1	4424334424444
to play	1	4444444444444
was going to play	10	2112211111231
was paying	1	4444444444444
was plaing	1	4444444444444
was playing	107	2121411111111
was to play	1	4121314221141
were playing	1	4444444444444
will be playing	2	4444444444444
will play	13	4444444444444
woold play	-	4444444444444
woud play	1	4444444424444
would be playing	21	2111112111241
would had been playing	1	4444444444444
would play	60	1121134421141
'll play / plays	1	-
was goig to play	1	-
woud be playing	1	-

Slot 55

had to walk	1	4221212423324
waked	-	4444444444444
walk	4	4444444444444
walked	312	1111111111111
went	2	4431444434444

Slot 56

had met	2	4434444444444
meat	1	4444444444444
meet	7	4444424444444
meeted	1	4444444444444
met	306	1111111111111
mett	1	4444444444444
meet(ed)	1	-

Slot 57

had reached	6	4434244444424
reach	4	4444424444444
reached	297	1111111111111
reachhed	-	4444444424444
reaching	2	4444444444444
reacht	1	4444444444444
was reach	1	4444444444444
was reaching	7	4424414444444
was reached	1	-

Slot 58

are looking	1	4444444444444
is looking	-	4444444444444
looced	-	4444444444444
look	1	4444444444444
looked	92	4444444444444
looks	1	4444444444444
was looking	217	1111111111111
was looking for	1	4414441421444
was lookking	1	4444444424444
was lookng	2	4444444444444
was woking	-	4444444444444
were looking	1	4444444444444
was / is looking	1	-
was lookig	1	-

Slot 59

call	2	4444424444444
called	294	1111111111111
did call	1	4234414441444
had a call	1	4444444424444
had called	1	4424444424434
was called	18	4444444444444
was calling	2	4444444444444

Slot 60

'd be	1	2444444444444
'll be	2	4444444444444
am	2	4444424444444
be	1	4444444444444
been	3	4444444444444
did be	1	4444444444444
had been	7	4444444444444
have	-	4444444444444
have been	5	4444444444444
was	250	4111111111111
was beeing	1	4444444444444
was being	5	4444244444444
was going to be	1	4434444424442
was going to be there	1	4434444424444
was there	2	4431414424434
were	11	4444444444444
were there	1	4444444444444
will be	1	4444444444444
would be	18	1424444444444
would have been	2	4424444444444
(no answer)	3	-
was (there)	1	-

Slot 61

'd planed	-	4444444444441
did plan	1	4444424444444
did planing	1	4444444444444
had plaint	-	4444444444444
had plan	1	4444444444444
had planed	14	4444444444444
had planned	124	1111111111111
hade planned	1	4444444444444

Slot 61 continues

have planed	-	4444444444444
have planned	2	4444424444444
plan	3	4444444444444
planed	18	4444444444444
plannd	-	4444444424444
planned	136	1424222123412
planted	2	4444444444444
was planed	3	4444444444444
was planing	2	4444444444444
was planned	3	4444444444444
was planning	3	4434444223434
(no answer)	2	-
(had) planned	1	-
had plant	1	-
plamed	1	-

Slot 62

had been visiting	2	4121221111211
had had to visit	1	4121211423224
had to visit	9	4431222124224
had visit	4	4444444444444
had visited	79	4111114121121
viseted	1	4444444444444
visit	3	4444444444444
visited	188	1224224224134
was going to visit	-	4444444444444
was visited	1	4444444444444
was visiting	23	4234324214414
were visiting	1	44444444447?4
would have visited	1	4444444444444
(no answer)	1	-
had to visit / visited	1	-
visiled	1	-
visited / had to visit	1	-
was goiing to visit	1	-
was visiting / visited	1	-

Slot 63

did let	1	4444444444444
had let	5	4444444444444
led	3	4434?44444444
left	-	4444444444444
let	300	1111111111111
lets	2	4444424444444
was letting	3	4444244424444
(no answer)	5	-

Slot 64

might seem	1	4444444444444
saw	2	4424444444444
seem	5	4444444444444
seemd	2	4444444444444
seemed	299	1111111111111
seemed to be	2	4311121422324
seems	3	4444424444444
seems to be	-	4444424444444
seemt	1	444444?444444

Slot 64 continues

was seeming	1	434444?444444
(no answer)	2	-
ceemed	1	-

Slot 65

shew	1	4444444444444
show	9	4444444444444
showd	2	4444444424444
showed	300	1111111111111
showm	-	4444444444444
shown	3	4444444444444
was showing	1	4434444444444
(no answer)	3	-

Slot 66

had been shouted	1	4444444444444
shouiting	1	4444444444444
shout	25	4124144244432
shouted	14	4444434444444
shouting	251	1111111111111
shoutting	2	4444444444444
to shout	3	4444444444444
was	1	4444444444444
was shouted	-	4444444444444
was shouting	19	4422214121442
shoutig	1	-
was shouling	1	-

Slot 67

are having	7	4434224444442
did have	1	4424224444444
had	149	1421114111214
had had	39	4111312114221
had have	6	4444444444444
had some	1	4341414424244
has	3	4444444444444
have	35	4444314441444
have been	1	4444444444444
have had	28	4434344244242
having	5	4444444444444
was having	1	4444444444444
were having	42	1121211114141
where having	1	4444444444444

Slot 68

had stopped	-	4414444444434
stop	1	4444444444444
stoped	22	4444444444444
stopid	-	4444444444444
stopped	293	1111111111111
stopt	1	4444444444444
was stoped	1	4444444444444
slopped	1	-

Slot 69

did go	2	4434424444434
get	1	4444444444444
gick	1	4444444444444

Slot 69 continues

go	1	4444424444444
goes	1	4444444444444
going	1	4444444444444
gone	1	4444444444444
got	2	4444444444444
was going	1	4434444444444
went	308	1111111111111

Slot 70

already had leeved	1	4444444444444
already had left	3	4133444421441
already leaved	2	4444444444444
already leaves	1	4444444444444
already leaving	1	4444444444444
already left	13	4444444444344
had allready leaved	-	4444444444444
had allready left	1	4414444424444
had alreade leaved	1	4444444444444
had alreade left	1	4444444444444
had already leave	3	4444444444444
had already leaved	21	4444444444444
had already left	222	1111111111111
had alredy leaved	1	4444444444444
had alredy left	4	4414444424444
has already leave	2	4444444444444
has already leaved	6	4444444444444
has already leaven	-	4444444444444
has already left	11	4444424444444
has been already leave	1	4444444444444
have already leaved	1	4444444444444
have already left	2	4444444444444
have left	1	4444444444444
leaved already	1	4444444444444
left already	4	4444444444344
was allready left	1	4444444444444
was already leave	-	4444444444444
was already leaved	2	4444444444444
was already leaving	6	4444444444444
was already left	3	4444444444444
was already lived	1	4444444444444
had alleady leaved	1	-
had alread left	1	-

Slot 71

get	16	4444424444444
got	283	1111111111111
got in	-	2321244424444
gott	-	4444444444444
had gotten	1	4444444444444
was getting	-	4434444444444
went	17	4442444444444
will got	1	4444444444444
got (in)	1	-

Slot 72

explain	45	4444444444444
explained	8	4444444444444

Slot 72 continues

explaining	34	4442414111312
to exlpain	1	4444444424444
to explain	230	1111111111111
to explaind	1	4444444444444

Slot 73

had	2	4444444444444
had have	1	4444444444444
has	6	4444444444444
have	29	4444244444444
have to have	1	4444444444444
having	13	4424444444444
to have	260	1111111111111
(no answer)	4	-
to get	2	-
hav	1	-

Slot 74

did not let	63	1111114111322
did not let me	2	4414444421444
didn't	1	4444444444444
didn't let	213	2411114111332
didn't let me	10	4114444421444
didnt let	1	44?4444424444
doesen't let	-	4444444444444
doesn't let	2	4444424444444
don't let	1	4444444444444
not let	1	4444444444444
won't let	1	4444444444444
would not let	8	4121211111111
would not let me	1	4424441421444
wouldn't let	13	4121211111111
(no answer)	1	-
would / did not let	1	-

Slot 75

'am	-	4444444444444
am	61	4242414441442
be	1	4441444444444
had been	5	4444424444444
has been	1	4444444444444
have	1	4444444444444
have been	2	4444444444444
m	1	4444444444444
to be	1	4444444444444
war	-	4444444444444
was	229	1111111111111
was beeing	-	4444444444444
was being	3	4224444241444
were	4	4444444444444
will be	1	4444444444444
would be	3	4444444444444
(no answer)	3	-
a were	1	-
am / was	1	-
woud be	1	-

Slot 76

did not know	57	1111111111111
did not known	-	4444444444444
didn't	1	4444444444444
didn't knew	8	4444444444444
didn't know	174	2111111111111
didnt	1	44?4444444444
do not know	1	4444444444444
doen't know	1	4444444444444
does	1	4444444444444
does not know	11	4122224441442
does'n't know	-	4444444444444
doesn't	2	4444444444444
doesn't know	28	4122224441342
don't know	33	4444444444444
was not knowing	-	4444444444444
wouldn't know	1	-

Slot 77

had talked	2	4434314444444
is talking	23	4132424441442
is talkings	1	4444444444444
talk	8	4444444444444
talked	49	4441114444444
talked about	1	4434444444444
talking	15	4444444444444
talking about	2	4444444444444
talks	12	4442444444444
was	3	4444444444444
was talked	-	4444444444444
was talking	196	1111111111111
was talking about	5	4414441421444
were talking	1	4444444444444
are talking	1	-

Slot 78

had realised	1	4424444444444
realise	4	4444444444444
realised	283	1111111111111
realised that	-	4211111221244
realized	30	1111411411111
relised	-	4444444444444
realisied	1	-

Slot 79

argue	21	4444444444444
argued	8	4444444444444
argueeing	-	4444444444444
argueing	23	4444414424444
arguing	101	4111141111111
argureering	1	4444444444444
of arguing	-	4444444444444
to argue	158	1442144121131
to argued	-	4444444444444
to arque	2	4444444444444
arquing	2	-
arque	1	-
to argue / arguing	1	-
to arguing	1	-

Slot 80

did leave	1	4234424444444
leaft	1	4444444444444
leaved	23	4444444444444
left	289	1111111111111
levt	1	444444?444444
was leaving	2	4444444444444
(no answer)	1	-
I left	1	-

Slot 81

being seen	-	4444444444444
saw	5	4444444444444
sawing	7	4444444444444
see	2	4444444444444
seeing	300	1111111111111
seing	3	444444?444444
to see	-	4444444444444
(no answer)	2	-

Slot 82

are not	7	4442444441411
aren't	3	4444444441211
did not be	3	4444444444444
did'nt be	1	4444444444444
didn't be	6	4444444444444
didn't was	2	4444444444444
didn't were	4	4444444444444
had not been	3	4424244444224
hadn't been	5	4424244444224
have been not	1	4434444444444
haven't be	1	4444444444444
not are	1	4444444444444
wasen't	-	4444444444444
wasn't	12	4444434444444
wasn't been	1	4444444444444
were no been	1	4444444444444
were not	113	1111111111111
were not a	-	4444444444444
were not being	-	4444444444444
weren:t	-	4444444424444
weren't	143	4111111111111
werent	3	4414444444444
wern't	1	4444444424444
where not	1	4444444444444
(no answer)	3	-
werenot	2	-
was not	1	-
were'nt	1	-

Slot 83

being	2	4444444444424
is	239	1111411111111
ist	-	4444444444444
lies	-	4424414424444
was	71	4442144414444
were	3	4444444444444
(no answer)	4	-

Slot 84

‘ve lived	2	4414112414111
am living	1	4444444444444
had been living	1	4444444444444
had lived	6	4444444444444
has live	1	4444444444444
has lived	2	4444444444444
have been lived	2	4444444444444
have been living	27	1121111111244
have lied	-	4444444444444
have live	3	4444444444444
have lived	254	1111112111111
have living	-	4444444444444
lived	17	4444444444434
(no answer)	3	-

Slot 85

‘ve got	2	4441414414414
also have	1	4441411424234
do have	1	4441414444442
had	152	4111114111441
had have	1	4444444444444
has	2	4444444444444
havd	-	4444444444444
have	151	1444414214112
have had	3	4434444224444
used to have	2	4121111411431
(no answer)	3	-
have / had	1	-

Slot 86

bought	302	1111111111111
buy	1	4444444444444
buyd	1	4444444444444
buyed	6	4444444444444
have bought	2	4434444444444
will buy	2	4444444444444
(no answer)	3	-
bougth	1	-
had bought	1	-

Slot 87

earn	16	4444444444444
earned	2	4444444444444
earnig	1	4444444444444
earning	96	4122112111211
to ean	1	4444444444444
to earn	200	1111111111111
(no answer)	3	-

Slot 88

did go	2	4134444444444
get	4	4444444444444
gick	1	4444444444444
go	1	4444444414444
goes	1	4444444444444
gone	1	4444444444444
got	12	4444444444444
had gone	1	4424442411444

Slot 88 continues

have went	1	4444444444434
used to go	1	4444444424434
was going	6	4134411421434
wen’t	1	4444444444444
went	282	1111142411111
(no answer)	3	-
vent	2	-

Slot 89

did go	2	4434444444444
get	1	4444444444444
gick	1	4444444444444
go	3	4444444444444
goes	3	4441444424434
got	1	4444444444444
had gone	-	4424344444434
left	1	4434444444444
was going	3	4134444424444
wen’t	1	4444444444444
went	298	1111111111111
wents	-	4444444444444
(no answer)	4	-
vent	1	-

Slot 90

did lose	-	4234444444444
loose	6	4444444444444
loosed	4	4444444444444
lose	6	4444444444444
losed	26	4444444444444
lost	273	1111111111111
losted	1	4444444444444
lust	-	4444444444444
(no answer)	3	-

Slot 91

have been looking	1	4444444444444
looked	138	4424444444414
took a look	1	4434444444434
was locking	-	4444444444444
was looking	173	1111111111131
were looking	3	4244444444444
(no answer)	3	-

Slot 92

meat	1	4444444444444
meet	3	4444424444444
meeted	1	4444444444444
met	307	1111111111111
mete	1	4444444444444
mett	1	4444444444444
saw	1	4431444434444
was meeting	1	4444444444444
(no answer)	3	-

Slot 93

had walked	1	4434444444444
walked	141	4134444444414

Slot 93 continues

walked up	1	4434444444444
walking	-	4444444444444
was waking	1	4444444444444
was walking	172	1111111111131
(no answer)	3	-

Slot 94

did seem	1	4134224443424
seamed	1	4444444444444
seem	1	4444444444444
seemd	2	4444444424444
seemed	298	1111111111111
seemed to be	6	4121111421234
seems	3	4444424444444
seems to be	1	4444424444444
seemt	2	4444444444444
sem	-	4444444444444
showed to be	-	4444444444444
was seeming	1	4444444444444
(no answer)	3	-

Slot 95

had been	-	4444444444444
had been seeing	1	4444444444444
had saw	17	4444444444444
had sawn	7	4444444444444
had seen	218	1111111111111
has been saw	1	4444444444444
has saw	3	4444444444444
has seen	8	44444444444?4
have been seen	1	4444444444444
have saw	-	4444444444444
have seen	10	4444444444444
save	1	4444444444444
saw	38	4424424424434
saws	1	4444444444444
see	2	4444444444444
seen	2	4444444444444
sees	1	4444444444444
was saw	1	4444444444444
was seen	1	4444444444444
would had seen	1	4444444444444
would have seen	2	4444224444444
(no answer)	3	-

Slot 96

did not	1	4444444424424
did not had	3	4444444444444
did not have	3	4444444444444
didn't	6	4444424424324
didn't had	10	4444444444444
didn't had seen	1	4444444444444
didn't has	1	4444444444444
didn't have	23	4444444444444
didnt have	-	44444444?444
din't had	1	4444444444444
had not	79	1111112114111
had not had	3	4444444444444
had not seen	1	4423444441444

Slot 96 continues

had not seen it	1	4421112224424
hadn't	146	2114111114111
hadn't had	8	4444444444444
hadn't have	1	4444444444444
hadn't seen	1	4424244441444
hadn't seen it	1	4421212224444
hadnt	-	4444444424444
has not	4	4444444444444
hasen't	-	4444444444444
hasn't	2	4444444444444
hasn't had	2	4444444444444
hasn't have	1	4444444444444
have not had	-	4444444444444
haven't	9	4444444444444
haven't had	4	4444444444444
would not had	1	4444444444444
(no answer)	3	-
hadn't (seen)	1	-
not had	1	-
she hadn't	1	-

Slot 97

did go	1	4434424444444
get	3	4444444444444
gick	1	4444444444444
goes	1	4444444444444
gone	1	4444444444444
got	3	4444444444444
had gone	1	4424444444444
was going	6	4434444444444
wen	1	4444444444444
wennt	1	4444444444444
went	297	1111111111111
(no answer)	3	-

Slot 98

did find	1	4434424444444
find	3	4444444444444
finded	5	4444444444444
foud	1	4444444444444
found	300	1111111111111
founded	3	4444444444444
foundt	1	4444444444444
fount	1	4444444444444
was finding	1	4444444444444
(no answer)	3	-

Slot 99

am not	1	4444444444444
can be	1	4444444444444
can not be	6	4142414144444
can not have been	1	2121414124344
can't be	11	4121114121314
cannot be	3	4121114121314
could not	1	4434444444424
could not be	9	4132214121324
could not have been	2	2121211121124
couldn't be	8	4132214121324
couldn't have been	2	2131211121144

Slot 99 continues

didn't be	5	4444444444444
didn't was	1	4444444444444
does not be	-	4444444444444
had not been	1	4444444444444
hadn't been	1	4444444444444
has not been	1	4444344444444
has to be	1	4444444444444
hasn't been	1	4444444444444
is not	36	4144114121434
isn't	28	4144114121434
isnt	1	4444444444444
was not	56	1412114111331
was'nt	-	444444444?44
wasn't	127	2411114111341
wasn't be	-	4444444444444
wasn't been	1	4444444444444
wont be	1	4444444444444
wouldn't be	1	4442414444444
(no answer)	4	-
can't	2	-
couldn't	1	-
hasn't	1	-
is not / can't be	1	-
was int	1	-
wasnt	1	-
wasntt	1	-

Slot 100

did not even	3	44?1444444414
did not even get	1	4434444444444
did not even go	52	1114111111141
did not even to go	1	4444444444444
did not even went	3	4444444444444
did not go even	1	4442444444444
did'n even go	1	4444444444444
didn't even	1	4444444444414
didn't even go	201	2111111111141
didn't even gone	1	4444444444444
didn't even went	5	4444444444444
didn't even went to	-	44444?4444444
didn't ever go	1	4444444444444
didn't evnt go	-	4444444444444
didnt even go	2	4444444424444
didntn even went	1	4444444444444
dindn't even go	1	4444444444444
dinn't even go	1	4444444444444
ditn't even go	1	4444444444444
does not even go	-	4444444444444
doesen't even get	-	4444444444444
doesn't even go	2	4444444424444
don't go ween	-	4444444444444
dosent even go	-	4444444444444
even did not go	-	4444444444444
even didn't go	2	4444444444444
even didn't went	1	4444444444444
even doesn't went	1	4444444444444
even gone	1	4444444444444
even went	1	4444444444444
had not even go	1	4444444444444

Slot 100 continues

had not even gone	4	4321344444124
had not even got	1	4442444444444
hadn't even go	2	4444444444444
hadn't even gone	3	4321314444124
hadn't even went	3	4444444444444
has not even gone	1	4444444244434
has not even went	1	4444444444444
hasn't	-	4444444444444
hasn't even go	1	4444444444444
haven't even went	1	4444444444444
haven't event gone	1	4444444444444
never went	1	4331444433144
not even went	2	4444444444444
was even not	1	4444444444444
wasn't even going	1	4444444444444
won't even go	1	4444444424444
(no answer)	5	-
didn't even vent	1	-
didn't event go	1	-
don't go even	1	-
even didn't wen't	1	-

Slot 101

did find it out	-	4144314444444
did find out	1	4144214444444
find out	6	4444444444444
finded out	7	4444444444444
finds out	5	4444424444444
fond out	1	4444444444444
fonded out	1	4444444444444
fouded out	1	4444444444444
found it out	16	4121212444444
found out	209	1121111111411
found out it	3	4243244444444
found that out	5	4131142241444
founed out	-	4444444444444
had found	2	444444444?44
had found it out	3	424444444?44
had found out	33	1411112111141
had found that out	2	4421214221444
had founded out	1	4444444444444
has find out	1	4444444444444
have finded it out	1	4444444444444
have found it out	1	4444444444444
was find out that	-	4444444444444
was fouding	1	4444444444444
was found out	2	4444444444444
will be found out	1	4444444444444
will find out	5	4444444444444
would find out	2	4444444444444
(no answer)	3	-
founded out	2	-
did found out	1	-
foud out	1	-
found (it) out	1	-
had found (it) out	1	-

Slot 102

could mean	1	4124414423434
didn't mean	-	4444444444444
does mean	1	4441414444444
doesn't	1	4444444444444
doesn't mean	1	4444444444444
is meaning	1	4444444444444
mead	-	4444444444444
mean	5	4444444444444
meane	-	4444444444444
meaned	3	4444444444444
meaning	1	4444444444444
means	271	1111111111111
means that	1	4112114211124
meant	11	4444444124444
menang	-	4444444444444
ment	5	4444444444444
met	-	4444444444444
will mean	-	4441434444444
would mean	11	4121122121341
would meen	1	4444444444444
(no answer)	4	-
wuold mean	1	-

Slot 103

did leave	3	4134214444444
didn't leave	1	4444444434444
had le	1	4444444444444
had leaved	5	4444444444444
had left	34	14211441244?4
had to leave	8	2441444442441
has leave	1	4444444444444
has left	5	4444444144434
have left	2	4444444444444
leave	3	4444444444444
leaved	15	4444444444444
leaven	-	4444444444444
leaves	1	4444444444444
left	229	1111111121114
levt	1	4444444444444
must have left	2	4121111121231
should have left	1	4124214424444
would have left	1	4134214111432
(no answer)	4	-
has left / would have left	1	-
have leaved	1	-

Slot 104

'd leave	1	4121114444442
'll leave	1	4431214444444
are going to leave	1	4444444444444
can't leave	1	4444444444444
didn't leave	1	4444444444444
going to leave	1	4444444444444
gonna leave	-	4444444444444
had leaved	3	4444444444444
had leaving	1	4444444444444
had left	26	4434444414414
had levt	1	4444444444444

Slot 104 continues

has left	4	4444444444444
is going to leave	5	4434414444444
is going to leaving	-	4444444444444
is leaving	10	4434414444444
leave	3	4444444444444
leaved	5	4444444444444
leaves	3	4444444444444
left	44	4434444444414
wants to leave	1	44?1414444444
was going to	2	4411444444434
was going to leave	27	2111111111241
was leaving	70	1121414111131
was left	-	4444444444444
was planning to leave	1	4121211433234
was to leave	1	4444414444344
were leaving	-	4444444444444
will leave	29	4441414444434
would be leaving	-	4421214113334
would leave	69	4121114124441
(no answer)	5	-
to leave	1	-
woud leave	1	-
wouldn't leave	1	-

Slot 105

murder	36	4444444444444
murdered	5	4444444444444
murdering	1	4444444444444
to have murdered	1	4434444444444
to murde	1	4444444444444
to murded	2	4444444444444
to murden	-	4444444444444
to murder	267	1111111111111
(no answer)	4	-
to murdered	1	-
tomurder	1	-

Slot 106

can't have walked	1	4141114114441
cannot have walked	1	41?1114114441
could not have walked	4	2111111111111
could not walk	5	41?1414444444
couldn't have walked	7	2111111111111
couln't walk	1	4444114444444
did not	3	4441444444444
did not walk	52	4431414244412
didn't	3	4444444444444
didn't had walk	1	4444444444444
didn't walk	189	4431414244412
didn't walked	2	4444444444444
didnt walk	3	4444444444444
don't walk	2	4444444444444
dosen't walk	-	4444444444444
had not walked	3	1434244444444
hadn't walk	1	4444444444444
hadn't walked	8	2434244444444

Slot 106 continues

has not walked	1	4444444444444
hasn't walked	2	4444444444444
not walk	1	4444444444444
not walked	1	4444444444444
walked not	1	4444444444444
was not walking	1	4444444444444
wasn't	1	4444444444444
wasn't walk	-	4444444444444
wasn't walked	1	4444444444444
wasn't walking	10	4444444444444
weren't walking	2	4444444444444
would not have walked	1	4424444444444
wouldn't have walked	2	4424444444444
wouldn't walk	2	4444444444444
(no answer)	4	-
didn't wak	2	-
couldn't walk	1	-

Slot 107

didn't meet	2	4444444444444
had met	39	4424244244414
had mett	1	4444444444444
has met	2	4444444444444
meat	1	4444444444444
meet	5	4444444444444
meets	1	4444424444444
met	260	1111111111111
saw	1	4441444434444
was going to	1	4444444444444
was meeting	-	4434444424444
(no answer)	4	-
had med	1	-
not	1	-

Slot 108

had been shouting at	1	4444444444444
had still been shouted at	1	4134144444444
had still shout at	1	4444444444444
has still been shouted at	1	4444444444444
have still shouted	1	4444444444444
is still shout at	1	4444444444444
is still shouted at	2	4444444444444
should be shouted at	1	4444444444444
shouted at still	1	4444444444444
shouted still at	1	4444444444444
shouted stll at	1	4444444444444
still did shout at	1	4444444444444
still shoted at	1	4444444444444
still shout	1	4444444444444
still shout at	5	4444444444444
still shouted	11	4444444444444
still shouted at	21	4444444444344
still shouted by	1	4444444444444
still shouts at	1	4444444444444

Slot 108 continues

still was shouting at	3	4444444444444
was being shouted at	1	4231444424344
was still at shouted	-	4444444444444
was still beeing shouted at	2	4441444424444
was still being shout at	1	4444444444444
was still being shouted at	28	2111111111111
was still being shouting at	1	4444444444444
was still getting shouted at	1	4141441244244
was still shout at	6	4444444444444
was still shouted	5	4444444444444
was still shouted at	48	1431244444444
was still shouted up	-	4444444444444
was still shouthed	1	4444444444444
was still shoutig at	1	4444444444444
was still shoutin at	1	4444444444444
was still shouting	48	4444444444444
was still shouting at	102	4444444444444
was still shoutting at	1	4444444444444
was till shouting	-	4444444444444
was till shouting at	2	4444444444444
were still shouted at	1	4444444444444
were still shouting at	2	4444444444444
would have been still shouting at	1	4444444444444
would still be shouted at	1	4134444444444
(no answer)	5	-
was still shouthng	1	-
was still shouting a	1	-
was till shouted at	1	-

Slot 109

are telling	1	4444444444444
did tell	1	4444414444444
had told	1	4444444444444
has been telling	5	4111111111231
has told	12	4431124214212
has told us	1	4431124424314
have told	4	44444?4444444
is lying	1	4432444434444
is teling	1	4444444444444
is telling	183	1141111111111
is telling lies	-	4421441424444
tell	3	4444444444444
tells	41	444424444447?4
told	47	1444224244424
told us	1	4431224444224
was telling	9	4431214121242
(no answer)	4	-
had totd	1	-
have told / is telling lies	1	-

Slot 109 continues

is tellig	1	-
told / (tells)	1	-

Slot 110

‘ll make	4	2144244444444
are going to make	3	4444444444444
are making	1	4444444444444
can make	-	4434444444444
do	1	4444444444444
have made	-	4444444444444
have to made	-	4444444444444
have to make	4	4444444444444
made	49	4411111411111
make	200	1144134141114
maked	-	4444444444444
making	1	4444444444444
should make	2	4431334444434
to make	13	4444444444444
will be making	1	4444444444444
will make	32	2144444444444
(no answer)	6	-
(’ll) make	1	-
made / will make	1	-